(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
1 August 2002 (01.08.2002)

PCT

(10) International Publication Number
WO 02/059381 A2

(51) International Patent Classification[7]: C12Q 1/68, C07K 14/47, C12N 15/63, 15/10, A01K 67/027

(21) International Application Number: PCT/US02/00473

(22) International Filing Date: 7 January 2002 (07.01.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/260,080 6 January 2001 (06.01.2001) US

(71) Applicant (*for all designated States except US*): **THE GENERAL HOSPITAL CORPORATION** [US/US]; 55 Fruit Street, Boston, MA 02114 (US).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **SLAUGENHAUPT, Susan** [US/US]; 23 Thomas Street, Quincy, MA 02169 (US). **GUSELLA, James, F.** [US/US]; 7 Woodstock Drive, Framingham, MA 01701 (US).

(74) Agent: **SONNENFELD, Kenneth, H.**; Morgan & Finnegan, LLP, 345 Park Avenue, New York, NY 10154 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: GENE FOR IDENTIFYING INDIVIDUALS WITH FAMILIAL DYSAUTONOMIA

(57) Abstract: This invention relates to methods and compositions useful for detecting mutations which cause Familial Dysautonomia. Familial dysautonomia (FD; Riley-Day syndrome), an Ashkenazi Jewish disorder, is the best known and most frequent of a group of congenital sensory neuropathies and is characterized by widespread sensory and variable autonomic dysfunction. Previously, we mapped the FD gene, *DYS*, to a 0.5 cM region of chromosome 9q31 and showed that the ethnic bias is due to a founder effect, with >99.5% of disease alleles sharing a common ancestral haplotype. To investigate the molecular basis of FD, we sequenced the minimal candidate region and cloned and characterized its 5 genes. One of these, *IKBKAP*, harbors two mutations that can cause FD. The major haplotype mutation is located in the donor splice site of intron 20. This mutation can result in skipping of exon 20 in the mRNA from FD patients, although they continue to express varying levels of wild-type message in a tissue-specific manner. RNA isolated from patient lymphoblasts is primarily wild-type, whereas only the deleted message is seen in RNA isolated from brain. The mutation associated with the minor haplotype in four patients is a missense (R696P) mutation in exon 19 that is predicted to disrupt a potential phosphorylation site. Our findings indicate that almost all cases of FD are caused by an unusual splice defect that displays tissue-specific expression; and they also provide the basis for rapid carrier screening in the Ashkenazi Jewish population.

WO 02/059381 A2

GENE FOR IDENTIFYING INDIVIDUALS WITH FAMILIAL DYSAUTONOMIA

This application claims priority to provisional application Serial No. 60/260,080, the entirety of which is incorporated herein by reference.

This invention was made with government support under Grant Number NS36326 awarded by The National Institutes of Health. The U.S. government has certain rights in the invention.

FIELD OF THE INVENTION

This invention relates generally to the gene, and mutations thereto, that are responsible for the disease familial dysautonomia (FD). More particularly, the invention relates to the identification, isolation and cloning of the DNA sequence corresponding to the normal and mutant FD genes, as well as characterization of their transcripts and gene products. This invention also relates to genetic screening methods and kits for identifying FD mutant and wild-type alleles, and further relates to FD diagnosis, prenatal screening and diagnosis, and therapies of FD, including gene therapeutics and protein/antibody based therapeutics.

BACKGROUND OF THE INVENTION

Familial Dysautonomia (FD, Riley-Day Syndrome, Hereditary Sensory and Autonomic Neuropathy Type III) [OMIM 223900] is an autosomal recessive disorder present in 1 in 3,600 live births in the Ashkenazi Jewish population. This debilitating disorder is due to the poor development, survival, and progressive degeneration of the sensory and autonomic nervous system (Axelrod et al., 1974). FD was first described in 1949 based on five children who presented with defective lacrimation, excessive sweating, skin blotching, and hypertension (Riley et al., 1949). The following cardinal criteria have evolved for diagnosis of FD: absence of fungiform papillae on the tongue, absence of flare after injection of intradermal

histamine, decreased or absent deep tendon reflexes, absence of overflow emotional tears, and Ashkenazi Jewish descent (Axelrod and Pearson, 1984, Axelrod 1984).

The loss of neuronal function in FD has many repercussions, with patients displaying gastrointestinal dysfunction, abnormal respiratory responses to hypoxic and hypercarbic states, scoliosis, gastroesophageal reflux, vomiting crises, lack of overflow tears, inappropriate sweating, and postural hypotension (Riley et al.1949; Axelrod et al.1974, Axelrod 1996). Despite recent advances in the management of FD, the disorder is inevitably fatal with only 50% of patients reaching 30 years of age. The clinical features of FD are due to a genetic defect that causes a striking, progressive depletion of unmyelinated sensory and autonomic neurons (Pearson and Pytel 1978a; Pearson and Pytel 1978b; Pearson et al. 1978; Axelrod 1995). This neuronal deficiency begins during development, as extensive pathology is evident even in the youngest subjects. Fetal development and postnatal maintenance of dorsal root ganglion (DRG) neurons is abnormal, significantly decreasing their numbers and resulting in DRG of grossly reduced size. Slow progressive degeneration is evidenced by continued neuronal depletion with increasing age. In the autonomic nervous system, superior cervical sympathetic ganglia are also reduced in size due to a severe decrease in the neuronal population.

Previously, the FD gene, *DYS*, was mapped to an 11-cM region of chromosome 9q31 (Blumenfeld et al. 1993) which was then narrowed by haplotype analysis to <0.5cM or 471 kb (Blumenfeld et al. 1999). There is a single major haplotype that accounts for >99.5% of all FD chromosomes in the Ashkenazi Jewish (AJ) population. The recent identification of several single nucleotide polymorphisms (SNPs) in the candidate interval has allowed for further reduction of the candidate region to 177 kb by revealing a common core haplotype shared by the major and one previously described minor haplotype (Blumenfeld et al. 1999).

SUMMARY OF THE INVENTION

This invention relates to mutations in the *IKBKAP* gene which the inventors of this invention discovered and found to be associated with Familial Dysautonomia. The mutation associated with the major haplotype of FD is a base

pair mutation, wherein the thymine nucleotide located at bp 6 of intron 20 in the *IKBKAP gene* is replaced with a cytosine nucleotide (T C) (hereinafter "FD1 mutation"). The mutation associated with the minor haplotype is a base pair mutation wherein the guanine nucleotide at bp 2397 (bp 73 of exon 19) is replaced with a cysteine nucleotide (G C) (hereinafter "FD2 mutation" This base pair mutation causes an arginine to proline missense mutation (R696P) in the amino acid sequence of the IKBKAP gene that is predicted to disrupt a potential phosphorylation site

In accordance with one aspect of the present invention, there is provided an isolated nucleic acid comprising a nucleic acid sequence selected from the group consisting of:

nucleic acid sequences corresponding to the genomic sequence of the FD gene including introns and exons as shown in Figure 6;

nucleic acid sequences corresponding to the nucleic acid sequence of the FD gene as shown in Figure 6, wherein the thymine nucleotide at position 34,201 is replaced by a cytosine nucleotide;

nucleic acid sequences corresponding to the nucleic acid sequence of the FD gene as shown in Figure 6, wherein the guanine nucleotide at position 33,714 is replaced by a cytosine nucleotide;

nucleic acid sequences corresponding to the nucleic acid sequence of the FD gene as shown in Figure 6, wherein the thymine nucleotide at position 34,201 is replaced by a cytosine nucleotide and the guanine nucleotide at position 33,714 is replaced by a cytosine nucleotide;

nucleic acid sequences corresponding to the cDNA sequence including the coding seqeunce of the FD gene as shown in Figure 7;

nucleic acid sequences corresponding to the cDNA sequence shown in Figure 7, wherein the arginine at position 696 is replaced by a proline;

In accordance with another aspect of the present invention, there is provided a nucleic acid probe, comprising a nucleotide sequence corresponding to a portion of a nucleic acid as set forth in any one of the foregoing nucleic acid sequences

In accordance with another aspect of the present invention, there is provided a cloning vector comprising a coding sequence of a nucleic acid as set forth above and a replicon operative in a host cell for the vector.

In accordance with another aspect of the present invention, there is provided an expression vector comprising a coding sequence of a nucleic acid set forth above operably linked with a promoter sequence capable of directing expression of the coding sequence in host cells for the vector.

In accordance with another aspect of the present invention, there is provided host cells transformed with a vector as set forth above.

In accordance with another aspect of the present invention, there is provided a method of producing a mutant FD polypeptide comprising: transforming host cells with a vector capable of expressing a polypeptide from a nucleic acid sequence as set forth above; culturing the cells under conditions suitable for production of the polypeptide; and recovering the polypeptide.

In accordance with another aspect of the present invention, there is provided a peptide product selected from the group consisting of: a polypeptide having an amino acid sequence corresponding to the amino acid sequence shown in Figure 8; a polypeptide containing a mutation in the amino acid sequence shown in Figure 8, wherein the arginine at position 696 is replaced with a proline; a peptide comprising at least 6 amino acid residues corresponding to the amino acid sequence shown in Figure 8, and a peptide comprising at least 6 amino acid residues corresponding to a mutated form of the amino acid sequence shown in Figure 8. In one embodiment, the peptide is labeled. In another embodiment, the peptide is a fusion protein.

In accordance with another aspect of the present invention, there is provided a use of a peptide as set forth above as an immunogen for the production of antibodies. In one embodiment, there is provided an antibody produced in such application. In one embodiment, the antibody is labeled. In another embodiment, the antibody is bound to a solid support. In accordance with another aspect of the present invention, there is provided a method to determine the presence or absence of the familial dysautonomia (FD) gene mutation in an individual, comprising: isolating genomic DNA, cDNA, or RNA from a potential FD disease

carrier or patient; and assessing the DNA for the presence or absence of an FD-associated allele, wherein said FD-associated allele is the FD1 and/or FD2 mutation wherein, the absence of either FD-associated allele indicates the absence of the FD gene mutation in the genome of the individual and the presence of the allele indicates that the individual is either affected with FD or a heterozygote carrier.

In one embodiment, the assessing step is performed by a process which comprises subjecting the DNA to amplification using oligonucleotide primers flanking the FD1 mutation and the FD2 mutation. In another embodiment, the assessing step further comprises an allele-specific oligonucleotide hybridization assay.

In another embodiment, DNA is amplified using the following oligonucleotide primers: 5'- GCCAGTGTTTTTGCCTGAG – 3'; 5'- CGGATTGTCACTGTTGTGC- 3'; 5'- GACTGCTCTCATAGCATCGC- 3'. In another embodiment, the assessing step further comprises an allele-specific oligonucleotide hybridization assay. In another embodiment, the allele-specific oligonucleotide hybridization assay is accomplished using the following oligonucleotides: 5'- AAGTAAG(T/C)GCCATTG- 3' and 5'- GGTTCAC(G/C)GATTGTC. In yet another embodiment, neuronal tissue from an individual is screened for the presence of truncated IKBKAP mRNA or peptides, wherein the presence of said truncated mRNA or peptides indicates that said individual possesses the FD1 and/or FD2 mutation in the IKBKAP gene.

In accordance with another aspect of the present invention, there is provided an animal model for familial dysautonomia (FD), comprising a mammal possessing a mutant or knock-out or knock-in FD gene. In another emodiment, there is provided a method of producing a transgenic animal expressing a mutant IKAP mRNA comprising:

(a) introducing into an embryonal cell of an animal a promoter operably linked to the nucleotide sequence containing a mutation associated with FD;

(b) transplanting the transgenic embryonal target cell formed thereby into a recipient female parent; and

(c) identifying at least one offspring containing said nucleotide sequence in said offspring's genome.

In accordance with another aspect of the present invention, there is provided a method for screening potential therapeutic agents for activity, in connection with FD, comprising: providing a screening tool selected from the group consisting of a cell line, and a mammal containing or expressing a defective FD gene or gene product; contacting the screening tool with the potential therapeutic agent; and assaying the screening tool for an activity.

In accordance with another aspect of the present invention, there is provided a method for treating familial dysautonomia (FD) by gene therapy using recombinant DNA technology to deliver the normal form of the FD gene into patient cells or vectors which will supply the patient with gene product in vivo.

In another embodiment, there is provided a method for treating familial dysautonomia (FD), comprising: providing an antibody directed against an FD protein sequence or peptide product; and delivering the antibody to affected tissues or cells in a patient having FD.

In accordance with another aspect of the present invention, there is provided kits for carrying out the methods of the invention. These kits include nucleic acids, polypeptides and antibodies of the present invention. In another embodiment the kit for detecting FD mutations will also contain genetic tests for diagnosing additional genetic diseases, such as Canavan's disease, Tay-Sachs disease, Goucher disease, Cystic Fibrosis, Fanconi anemia, and Bloom syndrome.

It will be appreciated by a skilled worker in the art that the identification of the genetic defect in a genetic disease, coupled with the provision of the DNA sequences of both normal and disease-causing alleles, provides the full scope of diagnostic and therapeutic aspects of such an invention as can be envisaged using current technology.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1. Genomic structure of *IKBKAP*. The figure illustrates the orientation and placement of the 37 exons within a 68 kb genomic region of

chromosome 9q31. The primers used for analysis of the splice defect are indicated as 18F (exon 18), 19F (exon 19) and 23R (exon 23). Asterick indicates the locations of the two mutations identified; the mutation associated with the major AJ haplotype is located at bp 6 of intron 20, whereas the mutation association with the minor AJ haplotype is located at bp 73 of exon 19. The 4.8 and 5.9 designations at exon 37 indicate the lengths of the two *IKBKAP* messages that differ only in the length of their 3' UTRs.

Figures 2A-2C. Demonstration of mutations in *IKBKAP*. Figure 2A shows the antisense sequence of the T – C mutation (shown by arrows adjacent to the G and A lanes) at bp 6 of intron 20 that is associated with the major FD haplotype. Lanes 1 and 2 are FD patients homozygous for the major haplotype (homozygous GG), lane 3 is an FD patient heterozygous for the major haplotype and minor haplotype 2 (heterozygous GA), lane 4 is an FD patient heterozygous for the major haplotype and minor haplotype 3 (heterozygous GA), and lanes 5 and 6 are non-FD controls (homozygous AA). Figure 2b shows heterozygosity for the G – C mutation (shown by arrows adjacent to the G and C lanes) at bp 73 of exon 19. Lane 1 is an FD homozygous for the major haplotype (homozygous GG), lanes 2-4 are three patients heterozygous for the major haplotype and minor haplotype 2 (heterozygous GC), lane 5 is a patient heterozygous for the major haplotype and minor haplotype 3 (homozygous GG), and lane 6 is a non-FD control (homozygous GG). Figure 2c shows the sequence of the cDNA generated from the RT-PCR of a patient heterozygous for the major and minor 2 haplotypes. The arrow points to the heterozygous G-C mutation in exon 19. The boundary of exons 19 and 20 is also indicated, illustrating that this patient expresses wild-type message that includes exon 20, despite the presence of the major mutation on one allele.

Figures 3A-3B. Northern blot analysis of *IKBKAP*. Figure 3A is a human multiple tissue northern blot that was hybridized with *IKBKAP* exon 2 and shows the presence of two messages of 4.8 and 5.9 kb (northern blots hybridized with other *IKBKAP* probes yielded similar patterns). Figure 3b is a northern blot generated using mRNA isolated from lymphoblast cell lines: lanes 1, 2, and 5 FD patients homozygous for the major haplotype; lane 3 individual carrying two definitively

non-FD chromososomes, lane 4 FD patient heterozygous for the major haplotype and minor haplotype 2; lane 6 control brain RNA (Clontech). The level of expression of *IKBKAP* mRNA relative to β-actin mRNA is quite variable in lymphoblasts. We observed no consistent increase or decrease in mRNA levels between FD patients homozygous for the major haplotype, those heterozyous for the major haplotype and minor haplotype 2, and non-FD individuals.

Figures 4A-4B: RT-PCR analysis of the exon 20 region of *IKBKAP* showing expression of the wild-type message and protein in patients. Figure 4A was generated using primers 18F (exon 18) and 23R (exon 23). Lanes 1 and 2 are FD patients homozygous for the major haplotype, lane 3 is an FD patient heterozygous for the major haplotype and minor haplotype 2, lanes 4 and 5 are non-FD controls, lane 6 is a water control. Figure 4b is a western blot generated using cytoplasmic protein isolated from patient lymphoblast cell lines and detected with a carboxyl-terminal antibody. Lanes 2, 4, 6, and 8 are patients homozygous for the major haplotype, lanes 3, 5, 7, and 9 are non-FD controls, lane 1 is a patient heterozygous for the major and minor haplotype 3, and lane 10 is a patient heterozygous for the major and minor haplotype 2 and lane 11 is a Hela cell line sample.

Figure 5. RT-PCR analysis of the exon 20 region of *IKBKAP* showing variable expression of the mutant message in FD patients. The analysis was done using primers 19F (exon 19) and 23F (exon 23). Lanes 1 and 2, control fibroblasts; lanes 3, 4, and 5, FD fibroblasts homozygous for the major mutation; lanes 6 and 7 FD lymphoblasts homozygous for the major mutation, lanes 8 and 9 non-FD lymphoblasts, lane 10 FD patient brain stem, lane 11 FD patient temporal lobe (showing a faint 319 bp band and no 393 bp band), lane 12 water control. RT-PCR of control brain RNA (Clontech) showed only the 393 bp band (data not shown).

Figure 6. The genomic sequence for *IKBKAP*.

Figure 7- The cDNA sequence for *IKBKAP*

Figure 8- the amino acid sequence of the *IKBKAP* gene

Figure 9- Comparison of the amino acid sequence of Ikap across several species. Alignment of the amino acid sequence of Ikap (M_musculus) with that of

Homo sapiens (H_sapiens), *Drosophila melanogaster* (D_melanogaster), *Saccharomyces cerevisiae* (S_cervisiae), *Arabidopsis thaliana* (A_thaliana), and *Caenorhabditis elegans* (C_elegans).

Figure 10- Comparison of the Novel Mouse *Ikbkap* Gene with Multiple Species Homologs

Figure 11- Mouse *Ikbkap* Exon and Intron Boundaries

Figure 12- Comparison of the synthetic regions of mouse chromosome 4 (MMU4) and human chromosome 9 (HSA9q31). This diagram on the left shows the location of *Ikbkap* in relation to mapped and genetic markers (boldface). Distances are given in centimorgans. The positions of the homologous genes that map to human chromosome 9q31 are shown on the right.

DETAILED DESCRIPTION OF THE INVENTION

This invention relates to mutations in the *IKBKAP* gene, which the inventors of the instant application discovered are associated with Familial Dysautonomia. More specifically, the mutation associated with the major haplotype of FD is a T-C change located at bp 6 of intron 20 in the *IKBKAP gene* as shown in Figure 1. This mutation can result in skipping of exon 20 in the mRNA from FD patients, although they continue to express varying levels of wild-type message in a tissue specific manner. The mutation associated with the minor haplotype is a single G-C change at bp 2397 (bp 73 of exon 19) that causes an arginine to proline missense mutation (R696P) that is predicted to disrupt a potential phosphorylation site.

These findings have direct implications for understanding the clinical manifestations of FD, for preventing it and potentially for treating it. The IKAP protein produced from *IKBKAP* gene was originally isolated as part of a large interleukin-1-inducible IKK complex and described as a regulator of kinases involved in pro-inflammatory cytokine signaling (Cohen et al. 1998). However, a recent report questioned this conclusion, by reporting that cellular IKK complexes do not contain IKAP based on various protein-protein interaction and functional

assays. Rather, IKAP appears to be a member of a novel complex containing additional unidentified proteins of 100, 70, 45, and 39 kDa (Krappmann et al. 2000).

IKAP is homologous to the Elp1 protein of *S. cerevisiae*, which is encoded by the *IKI3* locus and is required for sensitivity to pGKL killer toxin. The human and yeast proteins exhibit 29% identity and 46% similarity over their entire lengths. Yeast Elp1 protein is part of the RNA polymerase II-associated elongator complex, which also contains Elp2, a WD-40 repeat protein, and Elp3, a histone acetyltransferase (Otero et al. 1999). The human *ELP3* gene encodes a 60 kDa histone acetyltransferase that shows more than 75% identity with yeast Elp3 protein, but no 60 kDa protein has been found in the human IKAP-containing protein complex. Consequently, it is considered unlikely that IKAP is a member of a functionally conserved mammalian elongator complex (Krappmann et al. 2000). Instead, it has been reported that the protein may play a role in general gene activation mechanisms, as overexpression of IKAP interferes with the activity of both NF-κB-dependent and independent reporter genes (Krappmann et al. 2000). Therefore, the FD phenotype may be caused by aberrant expression of genes crucial to the development of the sensory and autonomic nervous systems, secondary to the loss of a functional IKAP protein in specific tissues.

FD is unique among Ashkenazi Jewish disorders in that one mutation accounts for > 99.5% of the disease chromosomes. As in other autosomal recessive diseases with no phenotype in heterozygous carriers, one might have expected to find several different types of mutations producing complete inactivation of the *DYS* gene in the AJ population. The fact that the major FD mutation does not produce complete inactivation, but rather allows variable tissue-specific expression of IKAP, may explain this lack of mutational diversity. Mutations causing complete inactivation of IKAP in all tissues might cause a more severe or even lethal phenotype. Indeed, *CG10535*, the apparent *Drosophila melanogaster* homologue of *IKBKAP*, maps coincident with a larval recessive lethal mutation (*l(3)04629*) supporting the essential nature of the protein (FlyBase). Thus, the array of mutations that can produce the FD phenotype may be limited if they must also allow expression of functional or partially functional IKAP in some tissues to permit

survival. With the identification of *IKBKAP* as *DYS*, it will now be possible to test this inactivation hypothesis in a mammalian model system.

Despite the overwhelming predominance of a single mutation in FD patients, the disease phenotype is remarkably variable both within and between families. The nature of the major FD mutation makes it tempting to consider that this phenotypic variability might relate to the frequency of exon 20 skipping in specific tissues and at specific developmental stages, which may be governed by variations in many factors involved in RNA splicing. Even a small amount of normal IKAP protein expressed in critical tissues might permit sufficient neuronal survival to alleviate the most severe phenotypes. This possibility is supported by the relatively mild phenotype associated with the presence of the R696P mutation, which is predicted to permit expression of an altered full-length IKAP protein that may retain some functional capacity. To date, this minor FD mutation has only been seen in four patients heterozygous for the major mutation. Consequently, it is uncertain whether homozygotes for the R696P mutation would display any phenotypic abnormality characteristic of FD. The single patient with minor haplotype 3 and mixed ancestry, whose mutation has yet to be found, is also a compound heterozygote with the major haplotype. The existence of minor haplotype 3 indicates that *IKBKAP* mutations will be found outside the AJ population, but like the R696P mutation, it is difficult to predict the severity of phenotype that would result from homozygosity.

Since FD affects the development and maintenance of the sensory and autonomic nervous systems, the identification of *IKBKAP* as the *DYS* gene allows for further investigation of the role of IKAP and associated proteins in the sensory and autonomic nervous systems. Of more immediate practical importance, however, the discovery of the single base mutation that characterizes >99.5% of FD chromosomes will permit efficient, inexpensive carrier testing in the AJ population, to guide reproductive choices and reduce the incidence of FD. The nature of the major mutation also offers some hope for new approaches to treatment of FD. Despite the presence of this mutation, lymphoblastoid cells from patients are capable of producing full-length wild-type mRNA and normal IKAP protein; while in neuronal tissue exon 20 is skipped, presumably leading to a truncated product.

Investigation of the mechanism that permits lymphoblasts to be relatively insensitive to the potential effect of the mutation on splicing may suggest strategies to prevent skipping of exon 20 in other cell types. An effective treatment to prevent the progressive neuronal loss of FD may be one aimed at facilitating the production of wild-type mRNA from the mutant gene rather than exogenous administration of the missing IKAP protein via gene therapy.

FD Screening

With knowledge of the primary mutation and secondary mutation of the FD gene as disclosed herein, screening for presymptomatic homozygotes, including prenatal diagnosis, and screening for heterozygous carriers can be readily carried out.

1. **Nucleic Acid Based Screening**

Individuals carrying mutations in the FD gene may be detected at either the DNA or RNA level using a variety of techniques that are well known in the art. The genomic DNA used for the diagnosis may be obtained from an individual's cells, such as those present in peripheral blood, urine, saliva, bucca, surgical specimen, and autopsy specimens. The DNA may be used directly or may be amplified enzymatically in vitro through use of PCR (Saiki et al. Science 239:487-491 (1988)) or other in vitro amplification methods such as the ligase chain reaction (LCR) (Wu and Wallace Genomics 4:560-569 (1989)), strand displacement amplification (SDA) (Walker et al. PNAS USA 89:392-396 (1992)), self-sustained sequence replication (3SR) (Fahy et al. PCR Methods Appl. 1:25-33 (1992)), prior to mutation analysis. in situ hybridization may also be used to detect the FD gene.

The methodology for preparing nucleic acids in a form that is suitable for mutation detection is well known in the art. For example, suitable probes for detecting a given mutation include the nucleotide sequence at the mutation site and encompass a sufficient number of nucleotides to provide a means of differentiating a normal from a mutant allele. Any probe or combination of probes capable of detecting any one of the FD mutations herein described are suitable for use in this

invention. Examples of suitable probes include those complementary to either the coding or noncoding strand of the DNA. Similarly, suitable PCR primers are complementary to sequences flanking the mutation site. Production of these primers and probes can be carried out in accordance with any one of the many routine methods, e.g., as disclosed in Sambrook et al..sup.45, and those disclosed in WO 93/06244 for assays for Goucher disease.

Probes for use with this invention should be long enough to specifically identify or amplify the relevant FD mutations with sufficient accuracy to be useful in evaluating the risk of an individual to be a carrier or having the FD disorder. In general, suitable probes and primers will comprise, preferably at a minimum, an oligomer of at least 16 nucleotides in length. Since calculations for mammalian genomes indicate that for an oligonucleotide 16 nucleotides in length, there is only one chance in ten that a typical cDNA library will fortuitously contain a sequence that exactly matches the sequence of the nucleotide. Therefore, suitable probes and primers are preferably 18 nucleotides long, which is the next larger oligonucleotide fully encoding an amino acid sequence (i.e., 6 amino acids in length).

By use of nucleotide and polypeptide sequences provided by this invention, safe, effective and accurate testing procedures are also made available to identify carriers of mutant alleles of *IKBKAP*, as well as pre- and postnatal diagnosis of fetuses and live born patients carrying either one or two mutant alleles. This affords potential parents the opportunity to make reproductive decisions prior to pregnancy, as well as afterwards, e.g., if chorionic villi sampling or amniocentesis is performed early in pregnancy. Thus, prospective parents who know that they are both carriers may wish to determine if their fetus will have the disease, and may wish to terminate such a pregnancy, or to provide the physician with the opportunity to begin treatment as soon as possible, including prenatally. In the case where such screening has not been performed, and therefore the carrier status of the patient is not known, and where FD disease is part of the differential diagnosis, the present invention also provides a method for making the diagnosis genetically.

Many versions of conventional genetic screening tests are known in the art. Several are disclosed in detail in WO 91/02796 for cystic fibrosis, in U.S. Pat. No.

5,217,865 for Tay-Sachs disease, in U.S. Pat. No. 5,227,292 for neurofibromatosis and in WO 93/06244 for Goucher disease. Thus, in accordance with the state of the art regarding assays for such genetic disorders, several types of assays are conventionally prepared using the nucleotides, polypeptides and antibodies of the present invention. For example: the detection of mutations in specific DNA sequences, such as the FD gene, can be accomplished by a variety of methods including, but not limited to, restriction-fragment-length-polymorphism detection based on allele-specific restriction-endonuclease cleavage (Kan and Dozy Lancet ii:910-912 (1978)), hybridization with allele-specific oligonucleotide probes (Wallace et al. Nucl Acids Res 6:3543-3557 (1978)), including immobilized oligonucleotides (Saiki et al. PNAS USA 86:6230-6234 (1989)) or oligonucleotide arrays (Maskos and Southern Nucl Acids Res 21:2269-2270 (1993)), allele-specific PCR (Newton et al. Nucl Acids Res 17:2503-25 16 (1989)), mismatch-repair detection (MRD) (Faham and Cox Genome Res 5:474-482 (1995)), binding of MutS protein (Wagner et al. Nucl Acids Res 23:3944-3948 (1995), denaturing-gradient gel electrophoresis (DGGE) (Fisher and Lerman et al. PNAS USA 80:1579-1583 (1983)), single-strand-conformation-polymorphism detection (Orita et al. Genomics 5:874-879 (1983)), RNAase cleavage at mismatched base-pairs (Myers et al. Science 230:1242 (1985)), chemical (Cotton et al. PNAS USA 85:4397-4401 (1988)) or enzymatic (Youil et al. PNAS USA 92:87-91 (1995)) cleavage of heteroduplex DNA, methods based on allele specific primer extension (Syvanen et al. Genomics 8:684-692 (1990)), genetic bit analysis (GBA) (Nikiforov et al. Nuci Acids Res 22:4167-4175 (1994)), the oligonucleotide-ligation assay (OLA) (Landegren et al. Science 241:1077 (1988)), the allele-specific ligation chain reaction (LCR) (Barrany PNAS USA 88:189-193 (1991)), gap-LCR (Abravaya et al. Nucl Acids Res 23:675-682 (1995)), and radioactive and/or fluorescent DNA sequencing using standard procedures well known in the art.

As will be appreciated, the mutation analysis may also be performed on samples of RNA by reverse transcription into cDNA therefrom. Furthermore, mutations may also be detected at the protein level using, for example, antibodies specific for the mutant and normal FD protein, respectively. It may also be possible

to base an FD mutation assay on altered cellular or subcellular localization of the mutant form of the FD protein.

2. **Antibodies**

Antibodies can also be used for the screening of the presence of the FD gene, the mutant FD gene, and the protein products therefrom. In addition, antibodies are useful in a variety of other contexts in accordance with this invention. As will be appreciated, antibodies can be raised against various epitopes of the FD protein. Such antibodies can be utilized for the diagnosis of FD and, in certain applications, targeting of affected tissues.

For example, antibodies can be used to detect truncated FD protein in neuronal cells, the detection of which indicates that an individual possesses a mutation in the IKBKAP gene.

Thus, in accordance with another aspect of the present invention a kit is provided that is suitable for use in screening and assaying for the presence of the FD gene by an immunoassay through use of an antibody which specifically binds to a gene product of the FD gene in combination with a reagent for detecting the binding of the antibody to the gene product.

Antibodies raised in accordance with the invention can also be utilized to provide extensive information on the characteristics of the protein and of the disease process and other valuable information which includes but is not limited to:

1. Antibodies can be used for the immunostaining of cells and tissues to determine the precise localization of the FD protein. Immunofluorescence and immuno-electron microscopy techniques which are well known in the art can be used for this purpose. Defects in the FD gene or in other genes which cause an altered localization of the FD protein are expected to be localizable by this method.

2. Antibodies to distinct isoforms of the FD protein (i.e., wild-type or mutant-specific antibodies) can be raised and used to detect the presence or absence of the wild-type or mutant gene products by immunoblotting (Western blotting) or other immunostaining methods. Such antibodies can also be utilized for therapeutic applications where, for example, binding to a mutant form of the FD protein reduces

the consequences of the mutation.

3. Antibodies can also be used as tools for affinity purification of FD protein. Methods such as immunoprecipitation or column chromatography using immobilized antibodies are well known in the art and are further described in Section (II)(B)(3), entitled "Protein Purification" herein.

4. Immunoprecipitation with specific antibodies is useful in characterizing the biochemical properties of the FD protein. Modifications of the FD protein (i.e., phosphorylation, glycosylation, ubiquitization, and the like) can be detected through use of this method. Immunoprecipitation and Western blotting are also useful for the identification of associating molecules that may be involved in the mammalian elongation complex.

5. Antibodies can also be utilized in connection with the isolation and characterization of tissues and cells which express FD protein. For example, FD protein expressing cells can be isolated from peripheral blood, bone marrow, liver, and other tissues, or from cultured cells by fluorescence activated cell sorting (FACS) Harlow et al., eds., Antibodies: A Laboratory Manual, pp. 394-395, Cold Spring Harbor Press, N.Y. (1988). Cells can be mixed with antibodies (primary antibodies) with or without conjugated dyes. If nonconjugated antibodies are used, a second dye-conjugated antibody (secondary antibody) which binds to the primary antibody can be added. This process allows the specific staining of cells or tissues which express the FD protein.

Antibodies against the FD protein are prepared by several methods which include, but are not limited to:

1. The potentially immunogenic domains of the protein are predicted from hydropathy and surface probability profiles. Then oligopeptides which span the predicted immunogenic sites are chemically synthesized. These oligopeptides can also be designed to contain the specific mutant amino acids to allow the detection of and discrimination between the mutant versus wild-type gene products. Rabbits or other animals are immunized with the synthesized oligopeptides coupled to a carrier such as KLH to produce anti-FD protein polyclonal antibodies. Alternatively, monoclonal antibodies can be produced against the synthesized oligopeptides using

conventional techniques that are well known in the art Harlow et al., eds., Antibodies: A Laboratory Manual, pp. 151-154, Cold Spring Harbor Press, N.Y. (1988). Both in vivo and in vitro immunization techniques can be used. For therapeutic applications, "humanized" monoclonal antibodies having human constant and variable regions are often preferred so as to minimize the immune response of a patient against the antibody. Such antibodies can be generated by immunizing transgenic animals which contain human immunoglobulin genes. See Jakobovits et al. Ann NY Acad Sci 764:525-535 (1995).

2. Antibodies can also be raised against expressed FD protein products from cells. Such expression products can include the full length expression product or parts or fragments thereof. Expression can be accomplished using conventional expression systems, such as bacterial, baculovirus, yeast, mammalian, and other overexpression systems using conventional recombinant DNA techniques. The proteins can be expressed as fusion proteins with a histidine tag, glutathione-S-transferase, or other moieties, or as nonfused proteins. Expressed proteins can be purified using conventional protein purification methods or affinity purification methods that are well known in the art. Purified proteins are used as immunogens to generate polyclonal or monoclonal antibodies using methods similar to those described above for the generation of antipeptide antibodies.

In each of the techniques described above, once hybridoma cell lines are prepared, monoclonal antibodies can be made through conventional techniques of, for example, priming mice with pristane and interperitoneally injecting such mice with the hybrid cells to enable harvesting of the monoclonal antibodies from ascites fluid.

In connection with synthetic and semi-synthetic antibodies, such terms are intended to cover antibody fragments, isotype switched antibodies, humanized antibodies (mouse-human, human-mouse, and the like), hybrids, antibodies having plural specificities, fully synthetic antibody-like molecules, and the like.

3. **Expression Systems**

Expression systems for the FD gene product allow for the study of the

function of the FD gene product, in either normal or wild-type form and/or mutated form. Such analyses are useful in providing insight into the disease causing process that is derived from mutations in the gene.

"Expression systems" refer to DNA sequences containing a desired coding sequence and control sequences in operable linkage, so that hosts transformed with these sequences are capable of producing the encoded proteins. In order to effect transformation, the expression system may be included on a vector; however, the relevant DNA may then also be integrated into the host chromosome.

In general terms, the production of a recombinant form of FD gene product typically involves the following:

First a DNA encoding the mature (used here to include all normal and mutant forms of the proteins) protein, the preprotein, or a fusion of the FD protein to an additional sequence cleavable under controlled conditions such as treatment with peptidase to give an active protein, is obtained. If the sequence is uninterrupted by introns it is suitable for expression in any host. If there are introns, expression is obtainable in mammalian or other eukaryotic systems capable of processing them. This sequence should be in excisable and recoverable form. The excised or recovered coding sequence is then placed in operable linkage with suitable control sequences in an expression vector. The construct is used to transform a suitable host, and the transformed host is cultured under selective conditions to effect the production of the recombinant FD protein. Optionally the FD protein is isolated from the medium or from the cells and purified as described in Section entitled "Protein Purification".

Each of the foregoing steps can be done in a variety of ways. For example, the desired coding sequences can be obtained by preparing suitable cDNA from cellular mRNA and manipulating the cDNA to obtain the complete sequence. Alternatively, genomic fragments may be obtained and used directly in appropriate hosts. The construction of expression vectors operable in a variety of hosts are made using appropriate replicons and control sequences, as set forth below. Suitable restriction sites can, if not normally available, be added to the ends of the coding sequence so as to provide an excisable gene to insert into these vectors.

The control sequences, expression vectors, and transformation methods are dependent on the type of host cell used to express the gene. Generally, prokaryotic, yeast, insect, or mammalian cells are presently useful as hosts. Prokaryotic hosts are in general the most efficient and convenient for the production of recombinant proteins. However, eukaryotic cells, and, in particular, yeast and mammalian cells, are often preferable because of their processing capacity and post-translational processing of human proteins.

Prokaryotes most frequently are represented by various strains of *E. coli*. However, other microbial strains may also be used, such as *Bacillus subtilis* and various species of Pseudomonas or other bacterial strains. In such prokaryotic systems, plasmid or bacteriophage vectors which contain origins of replication and control sequences compatible with the host are used. A wide variety of vectors for many prokaryotes are known (Maniatis et al. Molecular Cloning: A Laboratory Manual pp. 1.3-1.11, 2.3-2.125, 3.2-3.48, 2-4.64 (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1982)); Sambrook et al. Molecular Cloning: A Laboratory Manual pp. 1-54 (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1989)); Meth. Enzymology 68: 357-375 (1979); 101: 307-325 (1983); 152: 673-864 (1987) (Academic Press, Orlando, Fla. Pouwells et al. Cloning Vectors: A Laboratory Manual (Elsevier, Amsterdam (1987))). Commonly used prokaryotic control sequences which are defined herein to include promoters for transcription initiation, optionally with an operator, along with ribosome binding site sequences, include such commonly used promoters as the beta-lactamase (penicillinase) and lactose (lac) promoter systems, the tryptophan (trp) promoter system and the lambda derived PL promoter and N-gene ribosome binding, site, which has become useful as a portable control cassette (U.S. Pat. No. 4,711,845). However, any available promoter system compatible with prokaryotes can be used (Sambrook et al. supra. (1989); Meth. Enzymology supra. (1979, 1983, 1987); John et al . Gene 61: 207-215 (1987).

In addition to bacteria, eukaryotic microbes, such as yeast, may also be used as hosts. Laboratory strain Saccharomyces cerevisiae or Baker's yeast, is most often used although other strains are commonly available.

Vectors employing the 2 micron origin of replication and other plasmid vectors suitable for yeast expression are known (Sambrook et al. supra. (1989); Meth. Enzymology supra. (1979, 1983, 1987); John et al. supra. (1987). Control sequences for yeast vectors include promoters for the synthesis of glycolytic enzymes. Additional promoters known in the art include the promoters for 3-phosphoglycerate kinase, and those for other glycolytic enzymes, such as glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase. Other promoters, which have the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, and enzymes responsible for maltose and galactose utilization. See Sambrook et al. supra. (1989); Meth. Enzymology supra. John et al. supra. (1987). It is also believed that terminator sequences at the 3' end of the coding sequences are desirable. Such terminators are found in the 3' untranslated region following the coding sequences in yeast-derived genes. Many of the useful vectors contain control sequences derived from the enolase gene containing plasmid peno46 or the LEU2 gene obtained from Yepl3, however, any vector containing a yeast compatible promoter, origin of replication, and other control sequences is suitable (Sambrook et al. supra. (1989); Meth. Enzymology supra. (1979, 1983, 1987); John et al. supra.

It is also, of course, possible to express genes encoding polypeptides in eukaryotic host cell cultures derived from multicellular organisms (Kruse and Patterson Tissue Culture pp. 475-500 (Academic Press, Orlando (1973)); Meth. Enzymology 68: 357-375 (1979); Freshney Culture of Animal Cells; A Manual of Basic Techniques pp. 329-334 (2d ed., Alan R. Liss, N.Y. (1987))). Useful host cell lines include murine myelomas N51, VERO and HeT cells, SF9 or other insect cell lines, and Chinese hamster ovary (CHO) cells. Expression vectors for such cells ordinarily include promoters and control sequences compatible with mammalian cells such as, for example, the commonly used early and later promoters from

Simian Virus 40 (SV 40), or other viral promoters such as those from polyoma, adenovirus 2, bovine papilloma virus, or avian sarcoma viruses, herpes virus family (such as cytomegalovirus, herpes simplex virus, or Epstein-Barr virus), or immunoglobulin promoters and heat shock promoters (Sambrook et al. supra. pp. 16.3-16.74 (1989); Meth. Enzymology 152: 684-704 (1987); John et al. supra. In addition, regulated promoters, such as metallothionine (i.e., MT-1 and MT-2), glucocorticoid, or antibiotic gene "switches" can be used.

General aspects of mammalian cell host system transformations have been described by Axel (U.S. Pat. No. 4,399,216). Plant cells are also now available as hosts, and control sequences compatible with plant cells such as the nopaline synthase promoter and polyadenylation signal sequences are available (Pouwells et al. supra. (1987); Meth Enzymology 118: 627-639 (Academic Press, Orlando (1986); Gelvin et al. J. Bact. 172: 1600-1608.

Depending on the host cell used, transformation is done using standard techniques appropriate to such cells (Sambrook et al. supra. pp. 16.30-16.5 (1989); Meth. Enzymology supra 68:357-375 (1979); 101: 307-325 (1983); 152: 673-864 (1987). U.S. Pat. No. 4,399,216; Meth Enzymology supra 118: 627-639 (1986); Gelvin et al. J. Bact. 172: 1600-1608 (1990). Such techniques include, without limitation, calcium treatment employing calcium chloride for prokaryotes or other cells which contain substantial cell wall barriers; infection with Agrobacterium tumefaciens for certain plant cells; calcium phosphate precipitation, DEAE, lipid transfection systems (such as LIPOFECTIN.TM. and LIPOFFECTAMINE.TM.), and electroporation methods for mammalian cells without cell walls, and, microprojectile bombardment for many cells including, plant cells. In addition, DNA may be delivered by viral delivery systems such as retroviruses or the herpes family, adenoviruses, baculoviruses, or semliki forest virus, as appropriate for the species of cell line chosen.

C. THERAPEUTICS

Identification of the FD gene and its gene product also has therapeutic implications. Indeed, one of the major aims of this invention is the development of

therapies to circumvent or overcome the defect leading to FD disease. Envisioned are pharmacological, protein replacement, antibody therapy, and gene therapy approaches. In addition the development of animal models useful for developing therapies and for understanding the molecular mechanisms of FD disease are envisioned.

1. **Pharmacological**

In the pharmacological approach, drugs which circumvent or overcome the defective FD gene function are sought. In this approach, modulation of FD gene function can be accomplished by agents or drugs which are designed to interact with different aspects of the FD protein structure or function.

Efficacy of a drug or agent, can be identified in a screening program in which modulation is monitored in vitro cell systems. Indeed, the present invention provides for host cell systems which express various mutant FD proteins (especially the T-C and G-C mutations noted in this application) and are suited for use as primary screening systems.

In vivo testing of FD disease-modifying compounds is also required as a confirmation of activity observed in the in vitro assays. Animal models of FD disease are envisioned and discussed in the section entitled "Animal Models", below, in the present application.

Drugs can be designed to modulate FD gene and FD protein activity from knowledge of the structure and function correlations of FD protein and from knowledge of the specific defect in various FD mutant proteins. For this, rational drug design by use of X-ray crystallography, computer-aided molecular modeling (CAMM), quantitative or qualitative structure-activity relationship (QSAR), and similar technologies can further focus drug discovery efforts. Rational design allows prediction of protein or synthetic structures which can interact with and modify the FD protein activity. Such structures may be synthesized chemically or expressed in biological systems. This approach has been reviewed in Capsey et al., Genetically Engineered Human Therapeutic Drugs, Stockton Press, New York (1988). Further, combinatorial libraries can be designed, synthesized and used in

screening programs.

The present invention also envisions that the treatment of FD disease can take the form of modulation of another protein or step in the pathway in which the FD gene or its protein product participates in order to correct the physiological abnormality.

In order to administer therapeutic agents based on, or derived from, the present invention, it will be appreciated that suitable carriers, excipients, and other agents may be incorporated into the formulations to provide improved transfer, delivery, tolerance, and the like.

A multitude of appropriate formulations can be found in the formulary known to all pharmaceutical chemists: Remington's Pharmaceutical Sciences, (15th Edition, Mack Publishing Company, Easton, Pa. (1975)), particularly Chapter 87, by Blaug, Seymour, therein. These formulations include for example, powders, pastes, ointments, jelly, waxes, oils, lipids, anhydrous absorption bases, oil-in-water or water-in-oil emulsions, emulsions carbowax (polyethylene glycols of a variety of molecular weights), semi-solid gels, and semi-solid mixtures containing carbowax.

Any of the foregoing formulations may be appropriate in treatments and therapies in accordance with the present invention, provided that the active agent in the formulation is not inactivated by the formulation and the formulation is physiologically compatible.

2. Protein Replacement Therapy

The present invention also relates to the use of polypeptide or protein replacement therapy for those individuals determined to have a defective FD gene. Treatment of FD disease could be performed by replacing the defective FD protein with normal protein or its functional equivalent in therapeutic amounts.

FD polypeptide can be prepared for therapy by any of several conventional procedures. First, FD protein can be produced by cloning the FD cDNA into an appropriate expression vector, expressing the FD gene product from this vector in an in vitro expression system (cell-free or cell-based) and isolating the FD protein

from the medium or cells of the expression system. General expression vectors and systems are well known in the art. In addition, the invention envisions the potential need to express a stable form of the FD protein in order to obtain high yields and obtain a form readily amenable to intravenous administration. Stable high yield expression of proteins have been achieved through systems utilizing lipid-linked forms of proteins as described in Wettstein et al. J Exp Med 174:219-228 (1991) and Lin et al. Science 249:677-679 (1990).

FD protein can be prepared synthetically. Alternatively, the FD protein can be prepared from total protein samples by affinity chromatography. Sources would include tissues expressing normal FD protein, in vitro systems (outlined above), or synthetic materials. The affinity matrix would consist of antibodies (polyclonal or monoclonal) coupled to an inert matrix. In addition, various ligands which specifically interact with the FD protein could be immobilized on an inert matrix. General methods for preparation and use of affinity matrices are well known in the art.

Protein replacement therapy requires that FD protein be administered in an appropriate formulation. The FD protein can be formulated in conventional ways standard to the art for the administration of protein substances. Delivery may require packaging in lipid-containing vesicles (such as LIPOFECTIN.TM. or other cationic or anionic lipid or certain surfactant proteins) that facilitate incorporation into the cell membrane. The FD protein formulations can be delivered to affected tissues by different methods depending on the affected tissue.

3. **Gene Therapy**

Gene therapy utilizing recombinant DNA technology to deliver the normal form, of the FD gene into patient cells or vectors which will supply the patient with gene product in vivo is also contemplated within the scope of the present invention. In gene therapy of FD disease, a normal version of the FD gene is delivered to affected tissue(s) in a form and amount such that the correct gene is expressed and will prepare sufficient quantities of FD protein to reverse the effects of the mutated FD gene. Current approaches to gene therapy include viral vectors, cell-based

delivery systems and delivery agents. Further, ex vivo gene therapy could also be useful. In ex vivo gene therapy, cells (either autologous or otherwise) are transfected with the normal FD gene or a portion thereof and implanted or otherwise delivered into the patient. Such cells thereafter express the normal FD gene product in vivo and would be expected to assist a patient with FD disease in avoiding iron overload normally associated with FD disease. Ex vivo gene therapy is described in U.S. Pat. No. 5,399,346 to Anderson et al., the disclosure of which is hereby incorporated by reference in its entirety. Approaches to gene therapy are discussed below:

a. **Viral Vectors**

Retroviruses are often considered the preferred vector for somatic gene therapy. They provide high efficiency infection, stable integration and stable expression (Friedman, T. Progress Toward Human Gene Therapy. Science 244:1275 (1989)). The full length FD gene cDNA can be cloned into a retroviral vector driven by its endogenous promoter or from the retroviral LTR. Delivery of the virus could be accomplished by direct implantation of virus directly into the affected tissue.

Other delivery systems which can be utilized include adenovirus, adeno-associated virus (AAV), vaccinia virus, bovine papilloma virus or members of the herpes virus group such as Epstein-Barr virus. Viruses can be, and preferably are, replication deficient.

b. **Non-viral gene transfer**

Other methods of inserting the FD gene into the appropriate tissues may also be productive. Many of these agents, however, are of lower efficiency than viral vectors and would potentially require infection in vitro, selection of transfectants, and reimplantation. This would include calcium phosphate, DEAE dextran, electroporation, and protoplast fusion. A particularly attractive idea is the use of liposomes (i.e., LIPOFECTIN.TM.), which might be possible to carry out in vivo.

Synthetic cationic lipids and DNA conjugates also appear to show some promise and may increase the efficiency and ease of carrying out this approach.

4. **Animal Models**

The generation of a mouse or other animal model of FD disease is important for both an understanding the biology of the disease but also for testing of potential therapies.

The present invention envisions the creation of an animal model of FD disease by introduction of the FD disease causing mutations in a number of species including mice, rats, pigs, and primates.

Techniques for specifically inactivating or mutating genes by homologous recombination in embryonic stem cells (ES cells) have been described (Capecci Science 244:1288 (1989)). Animals with the inactivated homologous FD gene can then be used to introduce the mutant or normal human FD gene or for introduction of the homologous gene to that species and containing the T-C, G-C or other FD disease-causing mutations. Methods for these transgenic procedures are well known to those versed in the art and have been described by Murphy and Carter, Curr. Opin. Cell Biol. 4:273-279 (1992)

ILLUSTRATIVE EXAMPLES

The following examples are provided to illustrate certain aspects of the present invention and not intended as limiting the subject matter thereof.

Example 1

Identification of the *IKBKAP* gene and the mutations associated with FD were obtained as follows:

Patient Samples

Blood samples were collected from two major sources, the Dysautonomia Diagnostic and Treatment Center at New York University Medical Center and the Israeli Center for Familial Dysautonomia at Hadassah University Hospital, with

approval from the institutional review boards at these institutions, Massachusetts General Hospital and Harvard Medical School. Either F.A. or C.M. diagnosed all patients using established criteria. Epstein Barr virus transformed lymphoblast lines using standard conditions. Fibroblast cell lines were obtained from the Coriell Cell Repositories, Camden, NJ. RNA isolated from post-mortem FD brain was obtained from the Dysautonomia Diagnostic and Treatment Center at NYU. Genomic DNA, total RNA, and mRNA were prepared using commercial kits (Invitrogen and Molecular Research Center, Inc.). Cytoplasmic protein was extracted from lymphoblasts as previously described (Krappmann et al. 2000).

Identification of *IKBKAP* and mutation analysis

Exon trapping experiments of cosmids from a physical map of the candidate region yielded 5 exons that were used to screen a human frontal cortex cDNA library. Several cDNA clones were isolated and assembled into a novel transcript encoding a 1332 AA protein that was later identified as *IKBKAP* (Cohen et al. 1998). The complete 5.9 kb cDNA sequence of *IKBKAP* has been submitted to GenBank under accession number AF153419. In order to screen for mutations in FD patients, total lymphoblast RNA was reverse transcribed and overlapping sections of *IKBKAP* were amplified by PCR and sequenced. Evaluation of the splicing defect was performed using the following primers: 18F: GCCAGTGTTTTTGCCTGAG; 19F: CGGATTGTCACTGTTGTGC; 23R: GACTGCTCTCATAGCATCGC (Fig. 1).

DNA Sequencing

Sequencing was performed using the AmpliCycle sequencing kit (Applied Biosystems) or on an ABI 377 automated DNA sequencer using the BigDye terminator cycle sequencing kit (Applied Biosystems). The control sequence of the candidate region was obtained by constructing subclone libraries from BACs and sequencing using vector specific primers. The FD sequence was generated by sequencing cosmids from a patient homozygous for the major FD haplotype using sequence specific primers.

Expression Studies

Several human multiple tissue northern blots (Clontech) were hybridized using the following radioactively labeled probes: *IKBKAP* exon 2, *IKBKAP* exons 18/19/20, *IKBKAP* exon 23, and a 400 bp fragment of the *IKBKAP* 3'UTR immediately following the stop codon. Poly $(A)^+$ RNA was isolated from patient and control lymphoblast lines, northern blotted, and hybridized using a probe representing the full coding sequence of *IKBKAP*. Cytoplasmic protein extracted from lymphoblast cell lines was western blotted and detected using ECL (Amersham) with an antibody raised against a peptide comprising the extreme carboxyl terminus (AA 1313-1332) of human IKAP, the protein encoded by *IKBKAP* (Krappmann et al. 2000).

To identify *DYS*, exon trapping and cDNA selection were used to clone and characterize all of the genes in the 471 kb candidate region: *EPB41L8* (unpublished data) or *EHM2* (Shimizu et al. 2000), *C9ORF4* (Chadwick et al. 1999a), *C9ORF5* (Chadwick et al. 2000), *CTNNAL1* (Zhang et al.1998), a novel gene with homology to the glycine cleavage system H proteins (CG-8) (unpublished data), *IKBKAP* (Cohen et al. 1998), and *ACTL7A* and *ACTL7B* (Chadwick et al. 1999b). As FD is a recessive disorder, the *a priori* expectation for the mutation was inactivation of one of these genes. Consequently, each of these were screened for mutations by RT-PCR of patient lymphoblast RNA and direct sequencing of all coding regions. Although many SNPs were identified, there was no evidence for a homozygous inactivating mutation. Thus, it was concluded that the mutation would be found in non-coding sequence and the control genomic sequence of the entire 471 kb candidate region was generated using BACs from a physical map. Direct sequence prediction using GENSCAN and comprehensive searches of the public databases did not reveal any additional genes in the candidate region beyond those found by cloning methods. However, SNPs identified during sequence analysis enabled us to refine the haplotype analysis and narrow the candidate interval to 177 kb shared by the major haplotype and the previously described minor haplotype 1 (Blumenfeld et al. 1999). This reduced interval contains 5 genes, *CTNNAL1*, CG-8, *IKBKAP*, *ACTL7A* and *ACTL7B*, all previously screened by RT-PCR without yielding a

coding sequence mutation. A cosmid library was constructed from a patient homozygous for the major haplotype, assembled the minimal coverage contig for the now reduced candidate interval, and generated the sequence of the mutant chromosome.

Comparison of the FD and control sequences revealed 152 differences (excluding simple sequence repeat markers), which include 26 variations in the length of dT_n tracts, 1 VNTR, and 125 base pair changes. Each of the 125 base pair changes was tested in a panel of 50 individuals known to carry two non-FD chromosomes by segregation in FD families. Of the 125 changes tested, only 1 was unique to patients carrying the major FD haplotype. This T - C change is located at bp 6 of intron 20 in the *IKBKAP* gene depicted in Figure 1, and is demonstrated in Figure 2A. IKAP was originally identified as an IκB kinase (IKK) complex-associated protein that can bind both NF-κB inducing kinase (NIK) and IKKs through separate domains and assemble them into an active kinase complex (Cohen et al. 1998). Recent work, however, has shown that IKAP is not associated with IKKs and plays no specific role in cytokine-induced NF-κB signaling (Krappmann et al. 2000). Rather, IKAP was shown to be part of a novel multi-protein complex hypothesized to play a role in general transcriptional regulation.

The *IKBKAP* gene contains 37 exons and encodes a 1332 amino acid protein. The full-length 5.9 kb cDNA (GenBank accession number AF153419) covers 68 kb of genomic sequence, with the start methionine encoded in exon 2. *IKBKAP* was previously assigned to chromosome 9q34 (GenBank accession number AF044195), but it clearly maps within the FD candidate region of 9q31. Northern analysis of *IKBKAP* revealed two mRNAs of 4.8 and 5.9 kb (fig. 3a and b). The wild-type 4.8 kb mRNA has been reported previously (Cohen et al. 1998), while the second 5.9 kb message differs only in the length of the 3' UTR and is predicted to encode an identical 150 kDa protein. As seen in figure 3*b*, the putative FD mutation does not eliminate expression of the *IKBKAP* mRNA in patient lymphoblasts.

A base pair change at position 6 of the splice donor site might be expected to result in skipping of exon 20 (74 bp), causing a frameshift and therefore producing a

truncated protein. However, initial inspection of our RT-PCR experiments in patient lymphoblast RNA using primers located in exons 18 and 23 (Fig.1) showed a normal length 500 bp fragment that contained exon 20 (Fig. 4A), indicating that patient lymphoblasts express normal *IKBKAP* message. The Western blot shown in Figure 4B demonstrates that full-length IKAP protein is expressed in these patient lymphoblasts. However, as the antibody used was directed against the carboxyl-terminus of IKAP it would not be expected to detect any truncated protein should it be present. The presence of apparently normal IKAP in patient cells is at odds with the expectation of an inactivating mutation in this recessive disease.

In the absence of any evidence for a functional consequence of the intron 20 sequence change, the only alteration unique to FD chromosomes, additional genetic evidence was sought to support the view that it represents the FD mutation. The 658 FD chromosomes that carry the major haplotype all show the T – C change. *In toto*, 887 chromosomes have been tested that are definitively non-FD due to their failure to cause the disorder when present in individuals heterozygous for the major FD haplotype. None of these non-FD chromosomes exhibits the T – C mutation, strongly indicating that it is not a rare polymorphism. The frequency of the mutation in random AJ chromosomes was 14/1012 (gene frequency 1/72; carrier frequency 1/36), close to the expected carrier frequency of 1/32 (Maayan et al. 1987).

In view of the strong genetic evidence that this mutation must be pathogenic, it was postulated that its effect might be tissue-specific. RNA extracted from the brain stem and temporal lobe of a post-mortem FD brain sample was therefore examined. In contrast to FD lymphoblasts, RT-PCR of the FD brain tissue RNA using primers in exons 19 and 23 (expected to produce a normal product of 393 bp) revealed a 319 bp mutant product, indicating virtually complete absence of exon 20 from the *IKBKAP* mRNA (Fig. 5, lanes 10-11). As additional FD autopsy material could not be obtained, intensive analyses of additional lymphoblast and fibroblast cell lines were performed to determine whether exon-skipping could be detected. Fibroblast lines from homozygous FD patients yielded variable results. Some primary fibroblast lines displayed approximately equal expression of the mutant and

wild-type mRNAs while others displayed primarily wild-type mRNA. In addition, extensive examination of additional patient lymphoblast lines indicated that the mutant message could sometimes be detected at low levels. An example of the variability seen in FD fibroblasts and the presence of the mutant message in some FD lymphoblasts is shown in Figure 5. In fact, close re-examination of figure 4*a* shows a trace of the mutant band in 2 (lanes 1 and 2) of the 3 FD samples. The absence of exon 20 in the FD brain RNA and the preponderance of wild-type mRNA in fibroblasts and lymphoblasts indicate that the major FD mutation acts by altering splicing of *IKBKAP* in a tissue-specific manner.

To identify the mutations associated with minor haplotypes 2 and 3, (Blumenfeld et al. 1999) we amplified each *IKBKAP* exon, including adjacent intron sequence, from genomic DNA. A single G – C change at bp 2397 (bp 73 of exon 19) that causes an arginine to proline missense mutation (R696P) was identified in all 4 patients with minor haplotype 2 (fig. 2*b*). This was subsequently confirmed by RT-PCR in lymphoblast RNA as shown in figure 2*c* for a region that crosses the exon 19-20 border. The PCR product, generated from an FD patient who is a compound heterozygote with minor haplotype 2 and the major haplotype, clearly shows that RNA is being expressed equally from both alleles based on heterozygosity of the G – C point mutation in exon 19. However, the RNA from the major haplotype allele shows no evidence for skipping of exon 20 which would be expected to produce a mixture of exon 20 and 21 sequence beginning at the end of exon 19. This confirms our previous observation that lymphoblasts with the major FD mutation produce a predominance of normal *IKBKAP* transcript.

The R696P mutation is absent from 500 non-FD chromosomes, and it has been seen only once in 706 random AJ chromosomes in an individual who also carries the minor haplotype. This mutation is predicted to disrupt a potential threonine phosphorylation site at residue 699 identified by Netphos 2.0 (Blom et al. 1999), suggesting that it may affect regulation of IKAP. Interestingly, the presence of this minor mutation is associated with a relatively mild disease phenotype, suggesting that a partially functional IKAP protein may be expressed from this

allele. No mutation has been identified for minor haplotype 3, which represents the only non-AJ putative FD chromosome.

Example 2- FD Diagnostic Assays

As discussed above, the allele-specific oligonucleotide (ASO) hybridization assay is highly effective for detecting single nucleotide changes in DNA and RNA, such as the T-C or G-C mutations or sequence variations, especially when used in conjunction with allele-specific PCR amplification. Thus, in accordance with the present invention, there is provided an assay kit to detect mutations in the FD gene through use of a PCR/ASO hybridization assay.

PCR Amplification

Genomic DNA samples are placed into a reaction vessel(s) with appropriate primers, nucleotides, buffers, and salts and subjected to PCR amplification.

Suitable genomic DNA-containing samples from patients can be readily obtained and the DNA extracted therefrom using conventional techniques. For example, DNA can be isolated and prepared in accordance with the method described in Dracopoli, N. et al. eds. Current Protocols in Human Genetics pp. 7.1.1-7.1.7 (J. Wiley & Sons, New York (1994)), the disclosure of which is hereby incorporated by reference in its entirety. Most typically, a blood sample, a buccal swab, a hair follicle preparation, or a nasal aspirate is used as a source of cells to provide the DNA.

Alternatively, RNA from an individual (i.e., freshly transcribed or messenger RNA) can be easily utilized in accordance with the present invention for the detection of the FD2 mutation. Total RNA from an individual can be isolated according to the procedure outlined in Sambrook, J. et al. Molecular Cloning--A Laboratory Manual pp. 7.3-7.76 (2nd Ed., Cold Spring Harbor Laboratory Press, New York (1989)) the disclosure of which is hereby incorporated by reference.

In a preferred embodiment, the DNA-containing sample is a blood sample from a patient being screened for FD.

In amplification, a solution containing the DNA sample (obtained either directly or through reverse transcription of RNA) is mixed with an aliquot of each of dATP, dCTP, dGTP and dTTP (i.e., Pharmacia LKB Biotechnology, N.J.), an aliquot of each of the DNA specific PCR primers, an aliquot of Taq polymerase (i.e., Promega, Wis.), and an aliquot of PCR buffer, including MgCl.sub.2 (i.e., Promega) to a final volume. Followed by pre-denaturation (i.e., at 95.degree. C. for 7 minutes), PCR is carried out in a DNA thermal cycler (i.e., Perkin-Elmer Cetus, Conn.) with repetitive cycles of annealing, extension, and denaturation. As will be appreciated, such steps can be modified to optimize the PCR amplification for any particular reaction, however, exemplary conditions utilized include denaturation at 95.degree. C. for 1 minute, annealing at 55.degree. C. for 1 minute, and extension at 72.degree. C. for 4 minutes, respectively, for 30 cycles. Further details of the PCR technique can be found in Erlich, "PCR Technology," Stockton Press (1989) and U.S. Pat. No. 4,683,202, the disclosure of which is incorporated herein by reference.

In a preferred embodiment, the amplification primers used for detecting the T-C mutation and the G-C mutation in the FD gene are 5'-GCCAGTGTTTTTGCCTGAG – 3' / 5'-GACTGCTCTCATAGCATCGC- 3' and 5'- CGGATTGTCACTGTTGTGC- 3' / 5'-GACTGCTCTCATAGCATCGC- 3, respectively.

Hybridization

Following PCR amplification, the PCR products are subjected to a hybridization assay using allele-specific oligonucleotides. In a preferred embodiment, the allele-specific oligonucleotides used to detect the mutatons in the FD gene are as follows:

5'- AAGTAAG(T/C)GCCATTG- 3' and 5'- GGTTCAC(G/C)GATTGTC.

In the ASO assay, when carried out in microtiter plates, for example, one well is used for the determination of the presence of the normal allele and a second well is used for the determination of the presence of the mutated allele. Thus, the results for an individual who is heterozygous for the T-C mutation (i.e. a carrier of FD) will show a signal in each of the wells, an individual who is homozygous for

the T-C allele (i.e., affected with FD) will show a signal in only the C well, and an individual who does not have the FD mutation will show only one signal in the T well.

In another embodiment, a kit for detecting the FD mutation by ASO assay is provided. In the kit, amplification primers for DNA or RNA (or generally primers for amplifying a sequence of genomic DNA, reverse transcription products, complementary products) including the T-C mutated and normal alleles are provided. Allele-specific oligonucleotides are also preferably provided. The kit further includes separate reaction wells and reagents for detecting the presence of homozygosity or heterozygosity for the T-C mutation.

Within the same kit, or in separate kits, oligonucleotides for amplification and detection of other differences (such as the G-C mutation) can also be provided. If in the same kit as that used for detection of the T-C mutation, separate wells and reagents are provided, and homozygosity and heterozygosity can similarly be determined.

In another embodiment a kit combining other diseases (i.e., Canavan's)

Example 3- FD Diagnostic: Other Nucleotide Based Assays

As will be appreciated, a variety of other nucleotide based detection techniques are available for the detection of mutations in samples of RNA or DNA from patients. See, for example, the section, above, entitled "Nucleic Acid Based Screening." Any one or any combination of such techniques can be used in accordance with the invention for the design of a diagnostic device and method for the screening of samples of DNA or RNA for FD gene mutations in accordance with the invention, such as the mutations and sequence variants identified herein. Further, other techniques, currently available, or developed in the future, which allow for the specific detection of mutations and sequence variants in the FD gene are contemplated in accordance with the invention.

Through use of any such techniques, it will be appreciated that devices and methods can be readily developed by those of ordinary skill in art to rapidly and accurately screen for mutations and sequence variants in the FD gene in accordance

with the invention.

Thus, in accordance with the invention, there is provided a nucleic acid based test for FD gene mutations and sequence variants which comprises providing a sample of a patient's DNA or RNA and assessing the DNA or RNA for the presence of one or more FD gene mutations or sequence variants. Samples of patient DNA or RNA (or genomic, transcribed, reverse transcribed, and/or complementary sequences to the FD gene) can be readily obtained as described in Example 2. Through the identification and characterization of the FD gene as taught and disclosed in the present invention, one of ordinary skill in the art can readily identify the genomic, transcribed, reverse transcribed, and/or complementary sequences to the FD gene sequence in a sample and readily detect differences therein. Such differences in accordance with the present invention can be the T-C or G-C mutations or sequence variations identified and characterized in accordance herewith. Alternatively, other differences might similarly be detectable.

Kits for conducting and/or substantially automating the process of identification and detection of selected changes, as well as reagents utilized in connection therewith, are therefore envisioned in accordance with the invention of the present invention.

As discussed above, through knowledge of the gene-associated mutations responsible for FD disease, it is now possible to prepare transgenic animals as models of the FD disease. Such animals are useful in both understanding the mechanisms of FD disease as well as use in drug discovery efforts. The animals can be used in combination with cell-based or cell-free assays for drug screening programs.

EXAMPLE 4- Creating Animal Models of FD

The first step in creating an animal model of FD is the identification and cloning of homologs of the IKBKAP gene in other species.

Isolation of Mouse cDNA Clones

The human IKBKAP sequence (GenBank Accession No. AF153419) was used to search the mouse expressed sequence tag database (dbEST) using the BLAST

program (www.ncbi.nlm.nih.gov/BLAST). A single 5' EST from a mouse brain library (GenBank Association No. AU079160) was identified that showed marked similarity to the 5' end of IKBKAP. The corresponding cDNA clone, MNCB-3931, was obtained from the Japanese Collection of the Research Bioresource/National Institute of Infectious Disease. In addition, eight EST's that were similar to the 3' end of the ORF were found to belong to UniGene cluster Mn.46573 (www.ncbi.nlm.nih.gov/Unigene). Examination of this cluster yielded several poly (A+)- containing clones, and we obtained the clone UI-M-CG0p-bhb-g-07-0-U1 (GenBank Accession No. BE994893) from Research Genetics.

RT-PCR Analysis

RNA (1 ug/ml from BALB/c mouse brain was obtained commercially (Clontech). Oligo-dT 15 and random hexamer primers were annealed to the template at 65° C for 10 min in the presence of 1X first-strand buffer, 2mM dNTP mix, and 4 mM DTT. The reaction mixture was incubated at 42° C for 90 min after addition of Suuperscript TM II RT (200 U/ul) and Rnase inhibitor (80 U/ul) (GIBCO).

DNA Sequencing and Analysis

DNA sequencing was performed using the AmpliCycle sequencing kit (Applied Biosystems) for the 33 [P]-labeled dideoxynucleotide chain termination reaction, using the following conditions: 30 sec at 94° C, 30 sec at 60° C, and 30 sec at 72° C for 30 cycles. The radioactively labeled sequence reaction product was denatured at 95 C for 10 min and run on a denaturing 6% polyacrylamide gel for autoradiography. Basic sequencing manipulations and aligments were carried out using a program from Genetics Computer Group (GCC; Madison, WI). The cDNA sequence generated throughout the experiments were aligned and assembled into a 4799-bp cDNA named Ikbkap.

Isolation of Full-Length cDNA

To obtain the full-length cDNA sequence, PCR was performed on the mouse cDNA template using primers designed from the sequence of the 5' – and 3' –cDNA

clones. The PCR conditions were as follows: 15 sec at 95° C, 30 sec at 54° C to 60° C, and 3 min at 68° C for 9 cycles; then 15 sec at 95° C, 30 sec at 54 to 60° C, and 3 min with increment of 5 sec for each succeeding cycle at 68 C for 19 cycles, followed by 7 min at 72° C. The PCR products were electrophoresed on a 1% agarose gel stained with ethidium bromide and were cleaner using a Qiaquick PCR cleaning kit (Qiagen) in the preparation for cycle sequencing. Successive primers were designed in order to obtain the full-length Ikbkap sequence, which was deposited in GenBank under Accession No. AF367244.

Northern Blot Analysis

Expression of Ikbkap was examined using both mouse embryo and adult mouse multiple tissue Northern blots (Clontech). The blots were probed with a 1045-bp PCR fragment that contains exons 2 through 11, which was generated using primer 1 (5' –GGCGTCGTAGAAATTGC-3') and primer 2 (5' – GTGGTGCTGAAGGGGCAGGC-3'). The probe was radiolabeled (Sambrook et al., 1989) and was hybridized according to the manufacturer's instructions.

Chromosome Mapping of the Mouse Ikbkap Gene

Several of the mouse Ikbkap ESTs belogned to the Unigene cluster Mn.46573, which has been mapped to chromosome 4 (UniSTS entry: 253051) between D4Mit287 and D4Mit197. To assess synteny between mouse chromosome 4 and human chromosome 9, we used several resources available at NCBI (www.nbci.nlm.nih.gov/Homology).

Determination of Genomic Structure of the Mouse Ikbkap

The 37 human IKBKAP exons were searched against the Celera database to obtain homologous mouse sequences. Approximately 130 mouse genomic fragements (500-700 bp) were obtained using the Celera Discovery System and Celera's associated database, and these fragements were assembled into seven contigs. In order to assemble the coomplete genomic sequence, we obtaiined six mouse bacterial artificial chromosomes (BACs) from Researcg Genetics after they

screened an RPCI-23 mouse library using 4300bp human probe that contained exon 2. To verify that these BAC clones contained the entire Ikbkap gene, we amplified fragments from the 5' and 3' ends of the gene, as well as a fragment from the 3' flanking gene Actl7b (Slaugenhaupt et al., 2001) We designed primers at the ends of each of the seven contigs constructed from the Celera data and generated PCR products from the BACs. Subsequently, we sequenced and closed five of the gaps, with the resulting two contigs assembled and deposited to Celera (Accession No. CSN009).

Creating a Targeting Vector

After cloning and sequencing the mouse homolog of the human IKBKAP gene, a targeting vector can then be constructed from the mouse genomic DNA. The targeting vector would consist of two approximately 3 kb genomic fragments from the mouse FD gene as 5' and 3' homologous arms. These arms would be chosen to flank a region critical to the function of the FD gene product (for example, exon 20).

In place of exon 20, negative and positive selectable markers can be placed, for example, to abolish the activity of the FD gene. As a positive selectable marker a neo gene under control of phosphoglycerate kinase (pgk-1) promoter may be used and as a negative selectable marker the 5' arm of the vector can be flanked by a pgk-1 promoted herpes simplex thymidine kinase (HSV-TK) gene can be used.

The vector is then transfected into R1 ES cells and the transfectants are subjected to positive and negative selection (i.e., G418 and gancyclovir, respectively, where neo and HSV-TK are used). PCR is then used to screen for surviving colonies for the desired homologous recombination events. These are confirmed by Southern blot analysis.

Subsequently, several mutant clones are picked and injected into C57BL/6 blastocytes to produce high-percentage chimeric animals. The animals are then mated to C57BL/6 females. Heterozygous offspring are then mated to produce homozygous mutants. Such mutant offspring can then be tested for the FD gene mutation by Southern blot analysis. In addition, these animals are tested by RT-PCR

to assess whether the targeted homologous recombination results in the ablation of the FD gene mRNA. These results are confirmed by Northern blot analysis and RNase protection assays.

Once established, the FD gene-/-mice can be studied for the development of FD-like disease and can also be utilized to examine which cells and tissue-types are involved in the FD disease process. The animals can also be used to introduce the mutant or normal FD gene or for the introduction of the homologous gene to that species (i.e., mouse) and containing the T-C or G-C mutations, or other disease causing mutations. Methods for the above-described transgenic procedures are well known to those versed in the art and are described in detail by Murphy and Carter supra (1993).

The techniques described above, can also be used to introduce the T-C or G-C mutations, or other homologous mutations in the animal, into the homologous animal gene. As will be appreciated, similar techniques to those described above, can be utilized for the creation of many transgenic animal lines

To the extent that any reference (including books, articles, papers, patents, and patent applications) cited herein is not already incorporated by reference, they are hereby expressly incorporated by reference in their entirety.

While the invention has been described in connection with specific embodiments thereof, it will be understood that it is capable of further modification, and this application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure as come within known or customary practice in the art to which the invention pertains and as may be applied to the essential features hereinbefore set forth, and as fall within the scope of the invention and the limits of the appended claims.

WE CLAIM:

1. An isolated and purified nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of:

(a) the nucleic acid sequence in Figure 6;

(b) the nucleic acid sequence in Figure 6, wherein the thymine nucleotide at position 34,201 is replaced by a cytosine nucleotide;

(c) the nucleic acid sequence in Figure 6, wherein the guanine nucleotide at position 33,714 is replaced by a cytosine nucleotide;

(d) the nucleic acid sequence in Figure 6, wherein the thymine nucleotide at position 34,201 is replaced by a cytosine nucleotide and the guanine nucleotide at position 33,714 is replaced by a cytosine nucleotide;

(e) the nucleic acid sequence in Figure 7;

(f) the nucleic acid sequence in Figure 7, wherein the guanine nucleotide at position 2397 is replaced by a cytosine nucleotide.

2. An isolated and purified nucleic acid molecule according to claim 1, wherein the nucleic acid molecule has the nucleic acid sequence shown in Figure 6.

3. An isolated and purified nucleic acid molecule according to claim 1, said nucleic acid molecule having the nucleic acid sequence shown in Figure 6, wherein the thymine nucleotide at position 34,201 is replaced by a cytosine nucleotide.

4. An isolated and purified nucleic acid molecule according to claim 1, said nucleic acid molecule having the nucleic acid sequence shown in Figure 6, wherein the guanine nucleotide at position 33,714 is replaced by a cytosine nucleotide.

5. An isolated and purified nucleic acid molecule according to claim 1, said nucleic acid molecule having the nucleic acid sequence shown in Figure 6,

wherein the thymine nucleotide at position 34,201 is replaced by a cytosine nucleotide and the guanine nucleotide at position 33,714 is replaced by a cytosine nucleotide,

6. An isolated and purified nucleic acid molecule according to claim 1, wherein the nucleic acid molecule has the nucleic acid sequence shown in Figure 7.

7. An isolated and purified nucleic acid molecule according to claim 1, said nucleic acid molecule having the nucleic acid sequence shown in Figure 7, wherein the guanine nucleotide at position 2397 is replaced by a cytosine nucleotide.

8. An isolated polypeptide comprising the amino acid sequence in Figure 8.

9. An isolated polypeptide comprising the amino acid sequence in Figure 8, wherein the arginine at position 696 is replaced by a proline.

10. A recombinant vector comprising a nucleic acid molecule according to claim 1.

11. The recombinant vector according to claim 10, wherein said nucleic acid molecule is operably linked to an expression control sequence suitable for expression of said nucleic acid sequence in a host cell.

12. A host cell comprising the recombinant vector according to claim 11, wherein said host cell is selected from a group comprising a strain of E.coli, Pseudomonas, Bacillus subtilis, Bacillus stearothermophilus, or other bacilli, other bacteria, yeast, other fungi, insect cells, plant cells, or murine, bovine, porcine, human or other mammalian cells.

13. A method of producing a wild-type IKAP polypeptide, comprising:

(a) culturing a host cell transformed with a vector of claim 10 containing a DNA molecule encoding for a wild-type IKAP polypeptide in a cell culture medium under conditions whereby the IKAP polypeptide is expressed, and

(b) isolating the thus-produced wild-type IKAP polypeptide.

14. A method of producing a mutant IKAP polypeptide, comprising:

(a) culturing a host cell transformed with a vector of claim 10 containing a DNA molecule encoding a mutant IKAP polypeptide in a cell culture medium under conditions whereby the mutant IKAP polypeptide is expressed, and

(b) isolating the thus-produced mutant IKAP polypeptide.

15. A method of screening a subject to determine if said subject has a mutation associated with FD, comprising:

(a) providing a biological sample containing the DNA of the subject to be screened;

(b) detecting FD mutations in said biological sample.

16. The method according to claim 15, wherein the FD mutation is a T-C mutation at position 34,201 in the DNA sequence of Figure 6.

17. The method according to claim 15, wherein the FD mutation is a G-C mutation at position 33,714 in the DNA sequence of Figure 6.

18. The method according to claim 15, wherein the FD mutation is a T-C mutation at position 34,201 and a G-C mutation at position 33,714.

19. The method according to claim 15, wherein the FD mutation is detected by an allele-specific oligonucleotide hybridization assay.

20. The method according to claim 15, wherein the DNA from said biological sample is amplified using oligonucleotide primers flanking the mutation.

21. The method according to claim 20, wherein the DNA is amplified with oligonucleotide primers 18F and 23R.

22. The method according to claim 20, wherein the DNA is amplified with oligonucleotide primers 19F and 23R.

23. The method according to claim 20, wherein the DNA is amplified with oligonucleotide primers 18F, 19F and 23R.

24. The method according to claim 23, wherein the amplified DNA is screened for FD mutations using an allele-specific oligonucleotide hybridization assay.

25. The method according to claim 24, wherein the hybridization assay is accomplished using probes that span the T-C mutation at nucleotide position 34,201 in the IKBKAP gene.

26. The method according to claim 24, wherein the hybridization assay is accomplished using probes that span the G-C mutation at nucleotide position 33,714 in the IKBKAP gene.

27. The method according to claim 24, wherein the hybridization assay is accomplished using probes selected from the following sequences:

(a) 5'- AAGTAAG(T/C)GCCATTG- 3', and

(b) 5'- GGTTCAC(G/C)GATTGTC- 3'.

28. The method according to claim 15, wherein the FD mutation is detected by method selected from the group consisting of:

(a) restriction-fragment-length-polymorphism detection based on allele-specific restriction-endonuclease cleavage,

(b) hybridization with allele-specific oligonucleotide probes including immobilized oligonucleotides or oligonucleotide arrays,

(c) allele-specific PCR, mismatch-repair detection (MRD),

(d) binding of MutS protein,

(e) denaturing-gradient gel electrophoresis (DGGE),

(f) single-strand-conformation-polymorphism detection,

(g) RNAase cleavage at mismatched base-pairs,

(h) chemical or enzymatic cleavage of heteroduplex DNA,

(i) methods based on allele specific primer extension,

(j) genetic bit analysis (GBA),

(k) oligonucleotide-ligation assay (OLA),

(l) allele-specific ligation chain reaction (LCR)

(m) gap-LCR, and

(n) radioactive and/or fluorescent DNA sequencing.

29. A kit for assaying for the presence of an FD mutation in an individual comprising at least one oligonucleotide probe capable of detecting the FD1 mutation or the FD2 mutation.

30. A kit according to claim 29, further comprising primers capable of amplifying the region containing said mutations.

31. A kit according to claim 30, wherein said primers are 18F and 23R.

32. A kit according to claim 30, wherein said primers are 19F and 23R

33. A kit according to claim 30, wherein said primers are 18F, 19F and 23R.

34. A kit according to claim 29, further comprising an oligonucleotide probe which specifically hybridizes to one or more additional mutant or wild-type genes, wherein said additional gene codes for a protein associated with an additional genetic disease.

35. A kit according to claim 34, wherein the additional genetic disease is selected from the group comprising: Canavan's disease, Tay-Sachs disease, Goucher disease, Cystic Fibrosis, Fanconi anemia, and Bloom syndrome.

36. A method of detecting a FD mutation in a sample, comprising isolation of RNA from a tissue sample, amplifying the RNA using primers in exons 19 and 23; determining whether said sample contains a mutant product or a wild-type product, wherein the identification of a mutant product indicates the presence of an FD mutation in said sample.

37. The method according to claim 36, wherein said RNA is isolated from neuronal tissue.

38. A method of detecting a FD mutation in a sample, comprising the utilization of an antibody capable of detecting a truncated protein product that is indicative of FD.

39. A method of producing a transgenic animal expressing a mutant IKAP mRNA comprising:

(a) introducing into an embryonal cell of an animal a promoter operably linked to the nucleotide sequence containing a mutation associated with FD;

(b) transplanting the transgenic embryonal target cell formed thereby into a recipient female parent; and

(c) identifying at least one offspring containing said nucleotide sequence in said offspring's genome.

40. The method according to claim 39, wherein said mutation is the FD1 mutation.

41. The method according to claim 39, wherein said mutation is the FD2 mutation.

42. The method according to claim 15, further comprising a determination of whether said individual is homozygous or heterozygous for said mutation.

43. An oligonucleotide for detecting a mutation associated with FD, said oligonucleotide having a sequence selected from sequences which detect an FD mutation or bind to a region flanking said FD mutation.

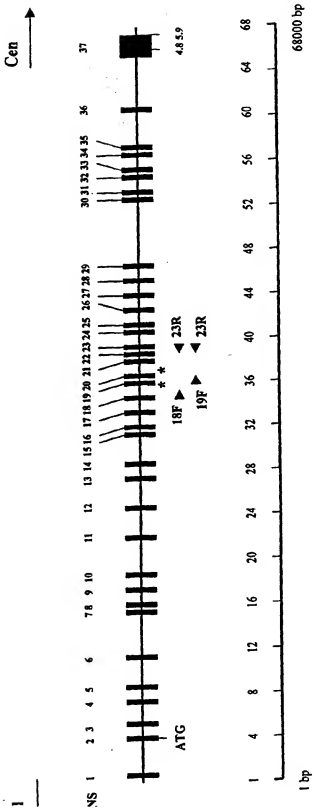


Fig. 1

Fig. 2a

Fig. 2b

Fig. 2c

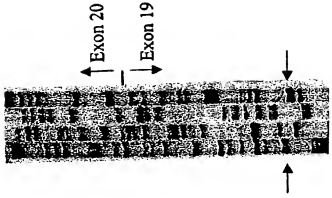


Fig. 3A

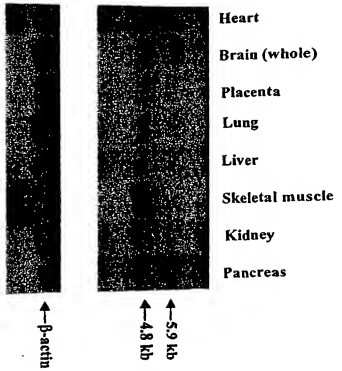


Fig. 3B

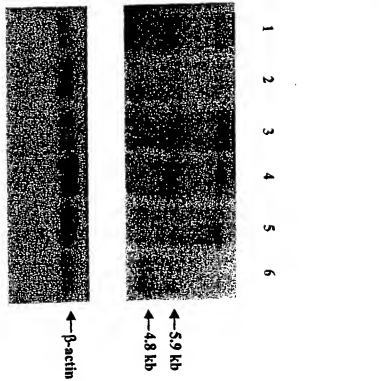


FIG. 4A

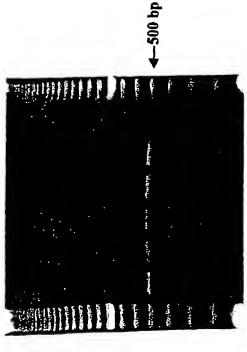
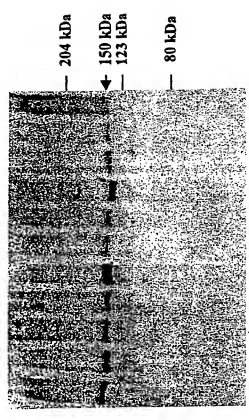


FIG. 4B



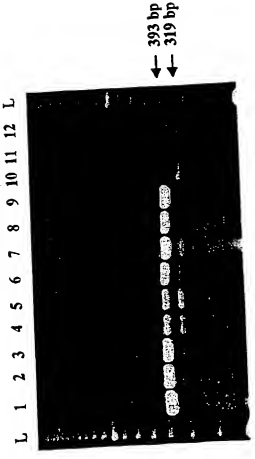


FIG. 5

FIGURE 6

IKBKAPgenomic.seq Length: 66479

```
   1 CCAGTGCTGC GGCTGCCTAG TTGACGCACC CATTGAGTCG CTGGCTTCTT
  51 TGCAGCGCTT CAGCGTTTTC CCCTGGAGGG CGCCTCCATC CTTGGAGGCC
 101 TAGTGCCGTC GGAGAGAGAG CGGGAGCCGC GGACAGAGAC GCGTGCGCAA
 151 TTCGGAGCCG ACTCTGGGTG CGGACTGTGG GAGCTGACTC TGGGTAGCCG
 201 GCTGCGCGTG GCTGGGGAGG CGAGGCCGGA CGCACCTCTG TTTGGGGGTC
 251 CTCAGGTAAG CGATCCATCC AGGGTAGGGG CACGGGAGTG GACCTCTCCG
 301 CCGGCGGTGT CCGGGTGAAG GAGACCCGGA GCCTCCTCTG CCTGCTGCGG
 351 GCCGGGGACT GGAGTGCGGG CTGCACCACC TCTTTCCTAG AGCCTTAAAT
 401 TCTTTTTGCA GCCTTGCCAC CTGCTCCATC GGGGGCGCTG GGAGGCGCGA
 451 CAGCCCAGGG ATGCCTGCTG CCCCTCCAGC CGGACTTAAC CCAGCCTCTT
 501 GATTGCTTGC AGGGGGTTGA TAATAACGCT GAAAGCGAGA GTATTAATTC
 551 ACGATGGAAG GCGGCGGTTA ATAGAGGCTC GGGTGCTGTG GTGCGGGTCC
 601 TTTCTCGCGT GTGAGACTTT TTCGTGGAGG TGGTGTCCTC TGTGCTTCTC
 651 CATCTAACGT GGTGTTTTAC GTGGCTTTCT CTCCCGTTAA CGATGATCTC
 701 CGTGGAGACA GTGGCTGAGT AATCTTCAGA TCCCAGTACT TAGCAAGTGC
 751 TCAGTCGGTG TTGGATGTAG GCCACAAACC GGATCGTAAA GAATTCAACT
 801 GTATATTGAC AGCCACGGAA CTAATCAATG AATAGATCCG TATGAAGAGT
 851 AAGCAAAAAG GCAGCAAAGA CAGTTTTTCA GCTTGGGGAC ATAGAGTAGA
 901 AATGGTCTGT CCCCAAATAG TGGGAACTGT CATTTGGGGG AAGAATAGCA
 951 AGTTCTTTGC TTTCCAGGTC GCATTTGATG TGCATGTGAG ACATGCTTGT
1001 GATTCTATCA GGAGGTTGAA AATGTGGGTT TAGTGGTAAG TTTGGGCTAA
1051 TTCAGTCAGG GCTAGGCATT TAGGCCTAAT CAGCGTATTG GTGATCTACC
1101 TGGTATATGT AATCATGCAT GTGATGTCTA GCCAAGAGGT GGATAGTCGA
1151 AGGAGCAAGG GAAGAAAATG AAGCAGTTAT CAGGAAATTA AGAGAGAATC
1201 CACGATTGAC CTTTGGTGTG GAGGGATCTT TAGCACATTT AAGAACTGCG
```

1251 AAGAGTTTGA ATCAGTGGAG GCAGGAAGGT TGGAGGTTGC AGATGTCCAA
1301 GAAAGAGTAC TAATAGGCCT AGGTCCTGTG GCAATATGGA GGATATTCCT
1351 TTCCTAGCCT GGAAAGAAGT GGAGGGAAGT CTTCCTCCGA GAAGATAAGG
1401 GAATAAGGCT GATGGGTGTG AAATTTCAGA GAAACTAGTT TTGAGGCGTT
1451 TTTATGATGT TTAAAGATGA AAAACGAGCA GGCACGGTGG CTCAGGCCTG
1501 TAATCCCAGC ACTTTGGGAG GCAGAGGCGG GTGGATCACT TGAGGTTAGG
1551 AGTTCAAGAA CAGCCTGGGC AACATGGTGA AACCCTGTCT CTACTAAAAA
1601 TaCAAAAaTT AACTGGGCAT GGTGCCGgGC GCCTGTAATC CCAGCTACTC
1651 CGGAGGCTGA GGCAGGAGAA TCGCTTGAAC CCGGAAGGCA GATGTTGCGG
17 TGAGCCGAGA TCGCGCCATT GCACCCCAGC CTGGGCAATA GAGCGAAAC
1 TCCGTCTCAA AAAACAAAAA AACCTGCATG ATATGTTAGA GGTTCAAGTA
1801 ATTTCTAGCA GTTCTTGAAT ATAATTGTCA CCAAAACTTA CTAAAATCAT
1851 TGTCTTCCTC ACTTCCATCA TATATAAACT TACCTTTCTC TTATCCCACA
1901 TTATATATTA TATAATTCCT ATGACACTTG ACATTATCTT CTGTGTACTA
1951 TTAGGATTGA TTCATCTTTA TTCTTTCTAT GTCATACATA TGTGGGGTGC
2001 CAAGATGAGA GAAGTCTCCT TGGATTAAAG TGACAATAAG ACCGGTGTGG
2051 TCCTTGTAAT TGCTACCCCT AACATAAGTT AGGGACTTAC AATCATAAGC
2101 CTTAAAGGGA TCTGAATATA AATAACTAGC ACAGTAACAT TTTTTTCCCC
2151 TACTTAGGTA ATGTTATGCA TTTAAGCAAG CCTGATTTTG CCAGACCAAA
2201 GTAGATGTCT TGTTTAGCAC TCTTTTCTCA CGTTTTATAT TGTCCTGGGA
2251 AAAGCCTGGC CAGAAGAACA AAGTTACTGG AAGTAGTTAT GTCAGGTCAT
2301 CAGGGTCCTT GAAATGTTGG TCATCATTTT GAAGTAAATT GTTGTCATGT
2351 CCCAGTATTT TCTCTTCCCC TTTAGAACAG TAAATGCTTT TCTATCTTTG
2401 ATTTCAGTTT TTTTATGAAT GTATAAAACC AGTTTATAAA TGAATAGACC
2451 TGGTGAATAT TAAAGTCATT TCAGATTCTC TTCAACTGCC AGTATTATAAA
2501 AATGGATTTT CAAATAGTGC TAATCAGTGG GATACCCTTT TGTTTTTCCT
2551 CATGATTTTA TAAAGATGTC CTAATATGCA AAAATAAAAT GTTTCCCCAT
2601 TCATTTGTTC TTTCAACTTT CCCAAAGGAA TAACTGATAT TACATCTTTT

594221_1

2651 TTGAAGAAAA CATTCTAAAG TTGAGAATCT TGCCTCTCCT AAAAAGAACA
2701 TAAAATAGGT TTCAGAATTC CTAATTTGTA GACCATAACT GTATAGAGTG
2751 GGTCAGGTTG CTGCTATAAT CCATACATGG GTGTGTACTC AGAGAGGTAA
2801 GTTTTTTCTT TTCTTGGTTA TTCTGATTCT GACTACCACT TCTTCACCCC
2851 CTGAATCATT TCATTTAAAT AAATATGGTC ATTTATCACT ATTAAGCTAT
2901 TTATTTTTCT CTTAGAGATT AATGATTCAT CAAGGGATAG TTGTACTTGT
2951 CTCGTGGGAA TCACTTCATC ATGCGAAATC TGAAATTATT TCGGACCCTG
3001 GAGTTCAGGG ATATTCAAGG TCCAGGGAAT CCTCAGTGCT TCTCTCTCCG
3051 AACTGAACAG GGGACGGTGC TCATTGGTTC AGAACATGGC CTGATAGAAG
3101 TAGACCCTGT CTCAAGAGAA GTAAGTTACT GATGTAGAAT GCCAGCATGT
3151 GGGTATGACC CTTGATTTCT CTTCTTCCAA ATTTCTTTCC CCACATGGTC
3201 TTTCTTTATA TCTTATTGAA TTTATATCCT CCCAAATAAA CATCTTTTGC
3251 TTCATATATA TGCCATGTTA GACATAGCTT AAATCGTAAT CCTTCTTTAA
3301 CTCTGCTGCT ATTTTAACCT AAGTCAGTAG AACTCTGACC TTACTTTTTG
3351 AGTGTGTGCC GTACTTTTTA CCCTCTTTGT CATGCAAATT CTGTTTATAA
3401 GAGTGGTTTT TTTTTTTTTT TTTTTGAGAC GGAGTCTCGC TCTGTCACCC
3451 AGGCTGGAGT GCAGTGGTGT GATCGTGGCT CACTGCAAGC TCCGCCTCCC
3501 CGGGTTCACA CCATTCTCCT GCCTCAGCCT CCCGAGAAGC TGGGACTACA
3551 GGCGCCCGCC ACCGCGCCCG GCTAATTTTT TGTATTTTTA GTAGATGTGC
3601 GGTTTCACCG TGTTAGCCAG GATGGTCTTG ATCTCCTGAC CTCGTGATCC
3651 GCCTGCCTCA GCGCCCGGCC AAGAGTGGTT TTTAATTGGG AATGAACACG
3701 AAAGTTGCCC ATGGAGCTTT CTAAAAGTTT GAGCCCACAT CTCATGTCAA
3751 CTAAATCAGA ATCTTTAGTG TTGGCTCCTA ACTATATGTA CTTTAAAAAC
3801 CTCTGTGGGT TGGTTTTGAT ATGGTCCCTT GATTATGTTC TTCTACTAAT
3851 ACATTTTAGG CAGTTACATC CTTTAGTGCC TTTTCCCCAT ACTATAGAAA
3901 TCTTAGAAAA GCATAGCTAT TAGCATCATA TTTTAGTGGA CAATTTTAAA
3951 GAGACCAGGC TTATTGTTTT TGTTTTTGTG TTTGTTTGGC AAAAAGGTCA
4001 CATTACCTAT TTTTCTTGTT AGAGATGACA GAGTAGTGAT ATTTCTCAAA

4051 TGAAAGTTTG GATTTTCATC TAGAAAAAAT ATTTTTGAAA GCTTTTATGT
4101 AATAAAAGAA GCATTAAAAA GTATTTCTGG AAATGTTATC AATTATTTCTT
4151 GAAAGTAGAC TGGGTTAATT TGCTTGTGTT TACTTTGGTG AAAGGTGAAA
4201 AATGAAGTTT CTTTGGTGGC AGAAGGCTTT CTCCCAGAGG ATGGAAGTGG
4251 CCGCATTGTT GGTGTTCAGG ACTTGCTGGA TCAGGAGTCT GTGTGTGTGG
4301 CCACAGCCTC TGGAGACGTC ATACTCTGCA GTCTCAGCAC ACAACAGGTA
4351 AGTGGAAGAC TCCAGTGAGG GGGGAGTCTC AAGCATCCTC AAATAGGTTA
4401 CTTGCTATTT GTGGAAGTTT TCAAATCAGT AGCCATAATA GTTACACTTT
4451 TGCTAATTAA TTTTTGCATT ATATATTTCT TTATTTAAAA AATTGTTAAC
4501 ATGGCTTTAT CTATATGTTA AGATTCTTCT AAAACTGAGT TTTGTCTGCT
4551 GCATCTATTA ATCAGAGTGA TCAGAATGTT CCAAATGAGA ATATATTTTT
4601 TTAAAAGTTA AAACTGGCTA TTCTTATGTG GTGTAGATCA CCTCTTATCA
4651 GACCCTCATC TTGAGTTGCA ACCTTTGTTT CTCAATTTAG GAAGTCTTTG
4701 TTTATCTGAC TTAGATTTTC TGTTATGAAT GTTGATTGGC TAAATTTAGA
4751 GTCCCTGAAG TCTAGGCACT AAAGTAAATA CATTGTCATT ACCTGCACAT
4801 GTGATGACTG CCAGTAGAGC TAGACTTCAA GCAATTGCTT CTTTCTCTAC
4851 TTTAGTGTAT AGTTGAGTTT CTGATTTCTA TCCTCACCTT CTTAACAGCA
4901 AGGGTTTCAA ATTACACTTG GCTGATTCTT TAAATCTTCT TCCATTACTT
4951 CATTAGTTGT GATCTCCTTA ACATTGATTA TGTCACAGAA GTTAGAGTAT
5001 TACTAATAGT AGGATAATGA TAGCAGCTTA CATTTATTAA CTATCATGTG
5051 CCTGGCACTT TTTAAAGTGC TTTTCATGCA AATTTATTTA ATCTTCACCA
5101 TGACCTTATG CAGTAGGTTG TTGTTTCCTA TTCTTCAGAA GAGGCAGTTA
5151 AGGCACAGAG TGCTTAAGTA ATTAGACCAG GGTCACACAG TAATCAAATG
5201 GGGTTTGACC CTAGCAGTCT AAATCTGGCA CCTCTGCTCT TAACCATTCC
5251 ATTTAGTACA ATCATAAACC TTTACTTGCA GTTCATGGTG GGAAATATCA
5301 AACTTGTCAT ATACAGCTTG TTTTTTTTTC GTATTTGAAA GATAGATGCT
5351 TTTACTTTCC AAACATTTTG TAGCATTGTT TCCTGGTTAC TGAGCTCTTC
5401 CAGTCTATTT ATCTTCATTT AATGGTGCTG ATTCTGCCCT TTAGTGGCTT

5451 CTCAATTGTC TGAAAGGTAG AGCCCACTAT TGTGCCTTAT AAGCCCCTTT
5501 CACTATCTGT TCCCCACATT CCTTTTTAGC CTCATCCCCC CATTGTTCCT
5551 GTGTGTACGT AAACCTTATG TTTTAGTTGC AGCTGATTTT TAACTGCTCT
5601 TTTTTCTGGC TTTGTGCCTC TACACTGTGT TTTCTTCCTG GTCTCTCTTT
5651 CCTGTCCTTA TTACCACTCT TTGAAACACG TCAGAAAAAC TTTTTCTGGA
5701 CTTTGGGCCA CTTGTCATTC CCTGTGCTGA GACGCATTTT GCTTTCCAGA
5751 GATCTTGGTC ATTGCTGTTA TCCTCTGTAG GGTCTTCTTT TATCTCCCTC
5801 GTGAGACAGC TCTGGGAAGA AAAAGATATT TATTTCTAAT CCCTGTGCCT
5851 AATAACAGGT CTATTCTCTT GATATCCATT ACTGAAGAAA TGTTTGTTGA
5901 GTAAGTTCTT GTTTTAATTT TTAAATATAA ATTTTTAATT TTTATGAGTA
5951 CATAGTAGGT ACATATATTT ATGGGCTACA TGAGATGTTC TGATACAGGC
6001 ATGCAGTGCA AAATAACCAC ATCATGGAGA ATAGGATATC CATCCCATCA
6051 AGCGTTTATC CTTTGTGTTA CAAACAATCC AATTACAGTC TTTTAGTTAT
6101 TTTAAAATGT GCAATTACTG TTGACTGTAG TTACCTTGTT GTGCTATCAA
6151 ATAGCAGGTC TTATTTATTC TATTTTTTTT TGTACCTATT AACCATCCCA
6201 ACTTCCCTCA GCCCCTCACT ACCCTTCCCA GCCTCTGGTA ACCATCCTTG
6251 TACTCTCTGT GTCCGTGAGT TCAATTGTTT TGATTTTTAG ATCGCACAAA
6301 TAAGTGAGAA CATGTGATGT TTGTCTTTTT GTGTCTGGTT TATTTCACTT
6351 AATGTAATCA TCTCCAGTTC CATCTATGTT GTTGCAGATG ACAGGATCTT
6401 ATTCTTTTTT TATGGCTGAA TAGTACTCCA TTGTGTATAG TACCACAATT
6451 TCTTTATCCA GTCATCCATT GATGGACACT TAGGTTGCTT CCAAATCTTA
6501 GCTATTGTGA ACAGAGCTGC AACAAACATG AGAGTGCAGA TATCTCTTCC
6551 ATATACTGAT GTCTTTTCGT TTTGTTTTTT TAATTGTTTT GATTGAAGTT
6601 GCAGTCAGTT TTTACTGAGA TGCTAGTGTT TGAATCTCTC TTTTCAATTT
6651 TCTCTGTCTC AGCTGGAGTG TGTTGGGAGT GTAGCCAGTG GTATCTCTGT
6701 TATGAGTTGG AGTCCTGACC AAGAGCTGGT GCTTCTTGCC ACAGGTAAGC
6751 TTGTTACTGG TGCCTCACTG GCTTTTTTAA AACATTATTC CAGATGTCTT
6801 ACAGGCTTCA TCAGCTTTAG GCTGCTTGAA TTTCAAAAAA TTTCTTTGAA

594221_1

6851 CCAGTATAAT ACCAATTATG AACCAGTATA ATACCAATTA TGTATGTGTG
6901 TGTGTATATA TATATAAAAC GTAGAGTGAT TTTTTTTTGG TGACTGAAGT
6951 TTTGCCTCTT AGTCTATCAT TATAAAAAGT TGTTTCATGT AACTTTTTAA
7001 GTCTTTGGGA GTAAGAAACA AAGTCATAAA ACTTGGGGAG GCTGCTAAGT
7051 CCCCAGTTAG AGTTAAAAAT GTCAGCAATA TGTATTTTAA CTTATTCTAA
7101 GAGTTGCTGT ATGGACACAT TCTAAAAGCC CTTCTTGGGT TCTGTTGCTG
7151 TTTTTCCCCT TTAAGTCTCA TCATTCCAGA TGAGTTTAGT AAACCAGCTC
7201 CACTGATGAC ATTTATATTT AGAGGTATCT TGGGGACAAG GAGTGTTGAA
7251 GTTAGTGGAG GAGGGCTTTG TGGACTTTTA AGTTCAACTG TACACACATT
7301 AATAGCTGAG CATAAGCACC AGGTGACTTA TCTAGGGAAA GCTTTTTGGG
7351 GTTTTTTGTC ATTGTTGTTT TTTTAAGTCA AAGCATTTTG GATGAATTCT
7401 GTCTGCTCTG TTCAGACTAA CTCCAGCTCC TTAGCTTACA GTGCCATAGG
7451 TACTTAGGAA TGGCAAATTT GTTACATGAA AACAAAATCA TTTTTGTTTG
7501 TGTTTCTCTA AGGTCAACAG ACCCTGATTA TGATGACAAA AGATTTTGAG
7551 CCAATCCTGG AGCAGCAGAT CCATCAGGAT GATTTTGGTG AAAGTAAGTA
7601 TAGCTTTGTG CAATATTTTG TGACCTACGT TTCTTCCCAT TTTTGACCAT
7651 TTCCTTGTGC ACTAATAGCC ATGTCATTAG GCCAAAGAAC TGTGAAAGTT
7701 AAACCCCCAG CTATTAAATG TCTATTAGCC CAGTTCCTTC AGCCCATCCC
7751 AAATCTTAAA AGGCCTACTG ATGCCTCTCC AGGTCTGAGG GTTTAAGGTC
7801 ACTTAGATAG TTATTACCCA AACCCTAGGA AAGTCTTAGG CTGCGGCTTTC
7851 AGTGAAAGGG ACTGTACAAG GTAGTATTTC TGGGATACAG TTTTAGGGAG
7901 AAGAAAAGAA GAAAGATGGA ATAGAAGGCT GGTTTTTGTT ACTACGATTA
7951 GATCCAATCT GCATTTCCAT GGGAACAATC AGATTATTTT CTTGCTAAAA
8001 TCTAGCCAAG GTCATCTGGG CATTAAGGCT GTGGGGGTAT TGAAGGGCAG
8051 TGCAGGAGAA GAGAGACGCT TATTAAGCAT AAGCTTTGGC CATCTTGAAG
8101 TCACAAAGTA GCTGGCCTGA TTGAAGAGGG ATGGGGAAGA AGATGTTCCA
8151 ACTTCTGTTA TGGTCTAACT TCCTGCCTTC TTGCTCCATC AACTCTGAGA
8201 AATCATTTAG ACAACTTCTA CCCATTTATT TACAAATAAT GTATTTGTTC

8251 AGAAATAATT TTGGAGGGCT GGGCACAGTG GCTCATGCCT GTAATCCCAG
8301 CACTTTTGGA GGATGAGGCA GGAGGATTGC TTGAGCCCAG GAGTTTGATA
8351 CTAGCCTAGG CAACGTAGGG AGACCCAGCA TCTACAAAGA ATTTAAAAAT
8401 TAGCTGGGCT TGGTGGTATC AGCACAGTAA TGACATGATG TGCAGGTACT
8451 GGGGTAGCAT AAGGGAAGGA AACGAGTAAC TAGAGAGGGA TGATTTATTT
8501 CCCCTAGGAG GCCAACTTGA GCTGAGTCTC AGCTGAATTG GTGTTGGGTA
8551 GGTGAGGGAT AAGGGTGGGG AGTAGTCAGC TGAATTGGTA TTGGGCAGGT
8601 GAGGGATAAG GGTTTGGAGT AGTCAGCTGA ATTGGTATTG GGTAGGTGAG
8651 GGATAAGGGT GGGGAACAGT CCAAGCAAGT GAATGTGTCC ATTTCAAGTG
8701 TCCATTTCAA GGGAGGGTTA TTTCATAGAA ACATTGTGGG TTACTCAGGG
8751 AACTGTGAGT AATTCAGCAT TGCTGAAGTG GCAGAATGTG AGTGTAGAAT
8801 GAAATAAATG GAACAGATTT GATTGAGTTT GTAGTAGGGA ATATGGACAT
8851 TGAGTTATAG TTGATCAGCC ATTACAAGTT TTGATGATAA GAGGTTTAAA
8901 GAGATTTATT TAATAGAAAG ATGGCTCGTG ATGGCATATT TTTGTTGTTT
8951 TTGTGTGTGG AGAGGGAAGA GATGAGAGGC AGGGTGATCA GGTAGGAGGT
9001 TGCTACAGGA ATCCAGATGA AAGATAAGGA AGGTTTGTGT GGGGCTAGAA
9051 GCAGGAATCA TTCAGGAAAA AACTTGATTC ACAATGAGGA TGGGAGTACA
9101 TTTTTTAGAA TTAGCTGGGA AACTTTTTTA GAATATATGT GCATGATTCC
9151 CCTTCTGCCC TAGGCCAGTT TGAGAAATAC CAATTTAGAA AGTGAAATAA
9201 ATAGGCTTTG CGTATGTAAG GTGAATAAGA AAAAGTTGAG CAGGACTCCA
9251 GCCAGAACCT CAGGTGTTGG GAATAAAGAT GCCAGTAACA GGGAAGATGG
9301 AGAAGTGCTG GTCTGTAAGG GGTGGGTGGT GAGATCTGTT TTGGATTTGT
9351 TGAAGGACCA TATGTGATTG CCATGTGGAG TATGCAAATA TAAGGCTGAA
9401 GCTCAGGAGA GGCCAGAGCT ATGGACTGAG AGTAGTGGGT ATGTAGGAAA
9451 TTCTGACAGT TTTGGGAACA GATGGACTGT CTCAGGGAGC AGATGCTGTA
9501 CAGGAAGAGT CTAGAATCCA GGGTGGAACT CTGGGGCATC CAGCTTTGAG
9551 GACAGTCAGA GAGAGAGTAA CAGCACACAG TATACTTTGG GATGGGAAAG
9601 TGCTCTGGGC CTGGTGTTTC CCACTGACTT TTTCACACAA ATCCTAATGC

9651 AGTAAATCAA AGGAAATGTA GGCCAAGTTA AGATCTTAGG TCTCAGAAAT
9701 GTGTTTCTCA GTACAAAAAA AAAAAAATCA TTCTATGGAG TGATGAATAT
9751 TTTTCCTCTA TCCTGGGGTC AGTAGACTTG TTCTGAAAAG GGCTAGGTCA
9801 TGAATATGTT CAGCTTTGCA GGCTGTATGA TCTGTGTTGC AGCTGCTCAA
9851 TTCTAATGTT GAGGTGTGAA AGTTATACAT GATACATAAG CACATCTATG
9901 TTCCAGTAAA CGTTTGTTTG TAAAAGCAGA TGTAGGCTGT AGTTTTGCAA
9951 ATCCCTGCTG TAACCCCATC ATTTCTTGTC TTCCATTGGA AAAGTTCTCT
10001 TTCTTCATTC CTTGGTCCTT AATCTTTCTG TGGAAACTTG CAGATAGAAG
10051 CCTGGGGGTT TGCACCAGGA TAGTCACTAC CATTTGTACG CAGCAGCAAT
10101 TGAGGTACTG TAGCACTTGG ATGTGAGCAG ACAGGAAATG GTCATATGGA
10151 CCCATAATTT ATAGGAATTG CAAACAGCCC TGCTTCATCA GAATCAGAAT
10201 CAATGGCAGG AGGAAAGTAT TGGGTCCTGG ATTAGGTGAT GTTTTCAGGA
10251 CCATCTTTAT TGTGCTTCTT GCAAATGGAT CCTACCTCCA GGAACAGAAG
10301 GGTTGTGTTG TTTCAGCAAC TCTGCCTAAT AGTTTATATA AGAGAAGTGT
10351 TACGATCTAG AAAGAACCCC AGTCAGCCTG GAAGGCAGAA GACCTGTGTT
10401 CTACTTTTTG GCTCCACCAT TAGGGAGGGT CTCAATCTCT AAGTCTATGT
10451 GAGGAGCTGT TTTGTGACCT GCAGCCCCTC TATCACCAGT GAGAGCTTGC
10501 AATCAGAATT TTATTCCCAG TTCTCATCTT GGGGTTTTAT GTTCCGGACA
10551 TATTTTGTAA ACTCTTTATG TTTCATTCTT CTTACTTATA AGGTGAGGGT
10601 GAGATCGCTG ACTTGTGTCA TCAAAGAAAC TTGGAATATG TAAGATGGCA
10651 GTAAAATGCT TTCCAAAATA AGGAAGGGCA TTTCAAATTC TTCAAAGTCA
10701 CTGCTGCATA TAATATGAAA TGGGTTTTGT TTGTTTGTTT TGAGATGGGG
10751 GTCTCGCTGT GTTACCCAGG CTAGAGAGTG CAGTAGTACA ATCAGGGCTC
10801 ACTGCAGCCT TGAACTCCTG GGTTCAAGTG ATCCTCCTAC TTTAGTCTCT
10851 TGAGTAGCTG GGACCACAGG TGTGTGCCAT CATGTCCAGC TTATTTTGTA
10901 TACTTTTTGT AGAGATGGGT GTCTCCCTAT GTTGCCCAGG CTGGTCTCGA
10951 ACTCCTGGAC TCAAGTGATC CTCCTGCCTC AGCCTCCCAA AGTGTTGGGA
11001 CTATAGGCAT GAGCCACCAT GCCCAGCCTG AAACATAGGT TTCTCAAATA

11051 TTGACTGCTG GTCAATTTAT TGAGAGGCGT TAGAGGACCT GAGTAATTGC
11101 CAATGACTAA CTTCATGAAG AATAGCAGTG AAACTGTTTT TGTTTCATTT
11151 CATGTGGCTT ATTAGTTGTC TTGCCAATTG TTCTGTAGGC AAGTTTATCA
11201 CTGTTGGATG GGGTAGGAAG GAGACACAGT TCCATGGATC AGAAGGCAGA
11251 CAAGCAGCTT TTCAGATGCA AATGGTAAGT TTGGTTTGAT GGATAAAAAG
11301 CCTTGACTGG AACAAATGTA AGTTTGCCAC CCACCAGGAA CTCTTTGGTG
11351 TCCACTTAGA TGCCAGTAAT GAACAGTTCT CTTCTGCTTT AGTAAAACTG
11401 CCTAGAACCT TCAGGAAATG AATCCCTCTA GAAAGATCCT TTTTTTCCTT
11451 GTTATTGCCA AGTTGCTTTG TGATTTATTT TCATAGTAGC AAATAATTAT
11501 AACCAATATT CATCACCCAG TTTAAAAAAT AAAACATCAC AGACAAAGGA
11551 AACCCCCTGT GTATCCCGTC CCGATGTCCC TCCCCTTCCT CTCCAGAGAG
11601 AGCTGCCATC CTTCATTCAC ATGCATGTTC TCATACTTTT CCCATATATG
11651 TGTATATTAG ATATTTTTCT TTTTCTGTTG GATGAAACTC TTTGTTTTCC
11701 TTACTTCTGG ATTGGAAAAT TCTGAAGACC ATATAATGAT GTCTTGATGA
11751 CTCAAGGCAG GACTTTTTAA TCTTCTAATG TAGGCGGGGC GGCCCCTGAA
11801 GGCAGAGGTG TGTGGACACA AGAAGAGTGC AGACTCTTGG GGCACCTGGG
11851 GAAGTAGTGT CCGTGTCACA TTAAATTCAT TTAAACTCTT ATATTTTATT
11901 TTAATTTATA CAATATGAAT ATTTTTTAAA ACTATGAATT GAAAAGTATT
11951 ACCCTTGAGT AAAATTAATG CCCCAAGAAG ATGTGCCATA TTTACCCTCT
12001 GGCACACTAC CAAGTACCCC CAGGGGCATT ACAGATCTCT GTTAGAAAAG
12051 TACAGATTAC ATTATCCTCA TAACATTTAG AAGCTATGAG ACCTTGGCAG
12101 GGAAGTTTCC TAATGTTTCT GAGCCTCAGT ATTCTCTGTA AAGTGGACAA
12151 CATAATGTCT CCTTACAAGG GTTGAGATGG GCAGGTAATA GCATATATAA
12201 AACAGCTATC ATAGCATCAG CACAGTGTAG GCACTCAAAT GGTAGTTGCT
12251 GCTTTTGTTT TAGTAGACAA ATAATTTTTG AAACTTTTTA AAGCGTAGTT
12301 TTTATTTCAA AACAACTTTA TTGTGAGTAA AATATGCATA GTGGGTCTAA
12351 TTTAACATTC TGAAAGCTAT TGACTTATTA GAACAGTAAA GGATTATTAG
12401 AGGGCAGAAA CATGGAGTAA GTACTCTGAG ACACAACCTT GCTTCTTTGG

12451 GGGTGATCCA CTACAACTGC CCAGCTTTGG ACAAGTGGTT TTCATGTTCC
12501 CCTGATTTTT AAGTGATTTT TTTTTTTTTT GGCAGGACTT AAAAGGTATC
12551 CTTGACTAAA CAGGAACTTG ACCAAGTAAA TAGTTGGTGC AATTTGAATA
12601 TTCTTTCTTG CTATAAGCAA CAAGTAAATT ATGGTACAGC TTTCTAAGAC
12651 CATATCTTTT CGATTTAAAA ATAGCACTTT ACTCATACAT GTTATGACAT
12701 GGGTAAACCT CATAAAGATT ATGCTAAGTG AAAGAAGCCA GTCATAAAAG
12751 ATCACATATA ATATGATCCC ATTTGTATGA AGTGCCCAGA AGGGGCAAAT
12801 CCACAGAGGC AGAAAGTAGA GTAGTGGTTG GGTAGGGCTG TGGGGTGGGG
12851 TGGGGAAGGG GTGACTGCTA ATGGATATGG GGTTTCTTTT GGGGATGATG
12901 AAAATGCTCA AAATTTAGAT TATGGTGATG GCTATTCAAC TTTGTAAATA
12951 TACTTTAAAA ACATTGATTC TTACCACTGA GTTTAAACAA CCAAAAAAAA
13001 ATCCCAAGGT GCATTGAATT GTGTACTTCA AATGGGTGAA CCTTAATAAT
13051 ATGTAAATTA TATCCCAGTA AAGGTGTTAA AAAATAGTAC TTTAAAGGAA
13101 TCTATGGTAG TTTTGAAAAT AAGGCAGTTT TCCATACTTT GTTAAACTCT
13151 GGAGAAGATG ACACTTTACT ACTGGTACCT GCTAGAGTAA GACTTATCTA
13201 GTATTAACAA AATTAGGGTT TATTAATGGT ATAGGATGAT CCAGGTAATG
13251 GGGGAAAAAA ACCGAGCATC CTGTTATCTA ATGTACTATC CAGTAAACTA
13301 CTCTAGCTTT TTTTCATGAA CTTTTTCTAA AGGCTTTCTA GGGCCTCGTC
13351 TTGGTTTGAA AGTTCACAGC TACCCTTCAG AAAAGAAAAC AAAAATCCAT
13401 GGAGTAGGCA GATACAAGTA CTCATGTGAG CATAATTTAC TTTGATTTTT
13451 TAAGTTGTGT TATTCTAGCC CTCAGCCTGT TCCCTGCCTG GGCTCTCCTA
13501 GTGCCCAGTA ACACTGATTC AAGAGGTTGC ATTTAGCTGG GCACAGTGGC
13551 TGATGCCTGC AATCCCAGCA CTTTGGGAGG CCAAGTTGGG CAGATCACCT
13601 GAGGTCAGGA GTTCAAGACC AGCATGTCCA ACATGGTGAA ATCCTATCTC
13651 TACTAAAAAT ACAAAAATTA GCCAGGCATG GTGGCAGATG CCTGTAATCT
13701 CAGCTACTTG AGAAGCTAAG GTAGTAGAAT CACTTGTACC TGGGAGGCAG
13751 AGGTTGCGGT GAGCCAAGAT TGTGCCACTG CACTCCAGCC TGGGCCATAA
13801 AGCAAGACTC CGTCTCAAAA AAAAAAAAAA AAAAATTGGG TGAGAGGGAG

13851 GAATTGAGGA GGATACCAAG GGTTGGGCCT GAACAAATGG AAGCATAATT
13901 ATATGTAGAA ATTTCTATGA GCTACTCTTC TAGAATAGAT GACTCAATAA
13951 TACCCTGCTT GCCATCTACG TTTTCTGTCC TTAATTATTT CCAGTTCTAT
14001 TTCATATAAT GCCTATTTCA GGCCTTAACC CTTCAGTAAA GGAGGTTTGG
14051 TTTCTATACC CTAGGACAGT TTCATTGAGA ATAAATTTTG TTAGGCTACC
14101 TATGTATTCC CTACTGTGCA GACTACAGTA CAGTACTAGC AGAATTCTTA
14151 GGCTGTTACT AGAATATGAT GATGAATGCC CGGGTGGTCA TCTGTCTCCC
14201 ACCCGGTAGA GTTGGCTTCA GGATTGAGAT ACACGTGGCC CTGGAGGAGA
14251 GTTTCTTCC CGTCATGCTG CAGAATGAGA ACATTTCCAT GTTTCGTCA
14301 TTGTCTGCTG CTGCCTTTAC CACCTCTGTG GCTCCTCCCT ATTCACCTTG
14351 TTCACATCTT ACTCATCTG TGCCCTGTTG TGAAGCTTAC ACAATATGTA
14401 AACAAAACTC TACCCTGTTG GACAAATGGA ACACTTGTTT CCTTGTTGTA
14451 GTTACCTGAT AGGTTCCTTA GCTCATTATA TTCAGGATCT AGATCTGTAG
14501 CTCTTTTCCT CTTTTGCTGT TCTCAGAGGC CACTTTTTTT TTTTTAATG
14551 CCGAAAGGAG GATTTTGTTT GTTTTACATT TTTTTCTTCT TTTTGATGAT
14601 TTCTGCGTTC TAAGAACCAA CCCTTGGATG GTTTCTGATT CTAGAGGCAG
14651 GCTTTCAAAG TAGCTTAAAC CTCTTAAAAA ACATCTGTAT CTAGTGGTCT
14701 GAGGCTTGTT TGATTCTGGG ATACTTAAGG TCCCCCAGTA ATATTGGTGT
14751 TTGTTCCCCT TTTTAGCATG AGTCTGCTTT GCCCTGGGAT GACCATAGAC
14801 CACAAGTTAC CTGGCGGGGG GATGGACAGT TTTTTGCTGT GAGTGTTGTT
14851 TGCCCAGAAA CAGGTATGGA AATATATTGC AGTTAAACAA CAATAAAAAA
14901 TTTTTATCTT ATTAAAATTA AGGAAAATTT TCTTTCTTTT GCTTTGAGTA
14951 GGGTATTAAT TATACATATG AGGCAAGGAT GTGCTGCTTT AAATGTGAAA
15001 TGAGGTTAGA GTTAAGAATT AGAAGAGTCC TTTGAGGCCA TTTGGTCCAT
15051 CCTCCTACCT GGTGGACACA AATTTGTAAC AAAATTAATC TAATTGGCTA
15101 TGTAAAACCA TGGCAGTTTT TATTTGTAAG GAAGGTGTTT GAATAGTTCT
15151 GAATTGACAA CTTTTATCAT AATGTTTTAA GTGTGTATGT GTGTTTGACT
15201 CCACTCCCGC ACAGGGGCTC GGAAGGTCAG AGTGTGGAAC CGAGAGTTTG

15251 CTTTGCAGTC AACCAGTGAG CCTGTGGCAG GACTGGGACC AGCCCTGGCT
15301 TGGAAGTGAG TGGGAGAAGA AACCTTAGAG AAATTCTTGG AACCAGAGTA
15351 GAGGTGGTGG TACACATGGA TACAGATGAT ACAGATGTTT GTGTAACACA
15401 AAAGGATTTT TACGTTTCTT CATTTGGTTA TAAGGCTGTA TCTATCTTTG
15451 TTTCTTCTTT TTTTTTTTTC TTATTCCCTG AAGTCTGAAT TCAACTCGAA
15501 TAGTAGATTT TACGCTTCTT CACAGATTTC ATTGTTCCAA GGCCGCATAT
15551 ATTTTGCATT CCTAACTCTT AAAAGGCTGT GGTTTTAAGG CAGGGTATAT
15601 ATGAAGCCAT TGTACAGAGC AGAAAATGGT GTTTAGAAGG GAAGGCCCAG
15651 TTTGCAAGGC TCTGTGGGGC AAATGGTGCT TTTGTGGAAA TTAGGGAAAG
15701 AGCCTCCTTC CTTGGCACAA AATTCCTACA GCAGAGGATC TGCTTGCCAA
15751 GGAGCATGCA GGCTGGATTC AGACCCTGCT CTTTCCTTCC ATTCTCCTCC
15801 TTGGCCCAGT ACCCTTGTGC AGGTTACAAT TTGCCTGTCA TATGTGGCTG
15851 CCTGATTTTA GATAGAAGAT GTATCTCCTC TGTTTCGGTG ATATCTGTTG
15901 TATGTAGACC TCTTGTTTCC CACCAGTATC TGAATGGTAT TATATGATAG
15951 AGCAGAAGAG AAATGTATTT GAATTAAAAC CCTAGAGACA AATATGAATA
16001 AGATGAGGCA ATTAAGATGT TTTCAACATT TGGTGAAGTC TTAAAAAAGA
16051 CCTACTGGAG CATAGAATAT TTGCTGAAGT TGTATAATGG AAGGAGAAAT
16101 AGATTTTGAT TTTTAGGACA TTATACCTGG AATGGTTTAG ATAACTTATT
16151 ATTTTTAAAG TCATCCAAAT GCAATGTAAA TATGTAAGGT TTTGTGGGCA
16201 AATGGAGCCT CTGTGTAAAA CAGGAAAAGG CACTCTTTCC TCTGGGCAAG
16251 TACAGTCCCA CAGTGGGATG AACCGCTCGC CGAGAGACAA GGGACACATG
16301 GGATTTAAAA CTTCCTTGGA TAAAGATATT CATTAATTCG TTCATTCATT
16351 CATTCATGTT TGCTGGAAAA AAAACTCTTC TGGATTTTAT CTATTCTTTA
16401 GTTAGGTGAG CTTTCGATAT TGTAACACTC TGAGTTTGCT TTAAGACCCT
16451 CAGGCAGTTT GATTGCATCT ACACAAGATA AACCCAACCA GCAGGATATT
16501 GTGTTTTTTG AGAAAAATGG ACTCCTTCAT GGACACTTTA CACTTCCCTT
16551 CCTTAAAGAT GAGGTTAAGG TAAGTGCCTG AGTTTGTTTC ACCCTCGAAT
16601 GTAGAGGACT TTCCATAGCT ATAGAGGGAA TTTTTTTTTT TTTTTTTTGA

16651 GATGGAGTTT CATTCTTGTT GCCCAGGTTG GAGTGCGATA GTGCAATCTC
16701 GGTTCACTGC AACCTCCGCC TCCTAGGTTC AAGTGATTCT CCTGCCTCAG
16751 CCTCCCGAGT AGCTGGGATT ACAGGCTTGC GCCACCACAG CCAGCTAATT
16801 TTGTATTTTT AGTAGAGACG GGGTTTCTCC GTGTTGGTCA GGCTGGTCTC
16851 AAACCCCTGA CCTCAGGTGA TCCACCCGCC TCTGCCTCCC AAAGTGCTGG
16901 GATTACAGGC GTGAGCCACC ACGCCTGGCC TATAGAGGGG ATTTATATTT
16951 GATATGGATA TATAAATAGT AGCTTTAGAG TAAATAGTAA TAAAATGGT
17001 GGCTTCCTAG AACTGATTTT TATTTAATAA AATATTGTTT TTCCAGTGAT
17051 TTTGCAAATA ATAGCATTTG TCCCCCACCT TAGATAAAAC AGAAGTAGGA
17101 AATAAAAATG CTAGTTTTTA TTGTTTATTT TGACAAAAGC ATAATTTTTC
17151 CAGTAATGAA GATGTTTTTC ATTTATAACA TTTAAATCTT AAGTGGTTTG
17201 TATACCATTA AGATTCTTGC TGAAGTGAGA ACACATCAAA TGGTATCTCT
17251 GTGTAAAATT TTAAACATCC TAAGTTGAGA GACGAGTTTA ATGAACTCCC
17301 ATGTAACTAT TACTCACTTT CAGTAGATAC CAACATTTTG CAAAACTATT
17351 TTCATCGGTC CGCAACTCTT TGGCCTATAC ATATATATAC TTACATATAT
17401 TTTTATTTCC TGGAGTTTTA ATTCTAGAAA TCATATTTTC AATATTTATT
17451 TATAACAGTT AAGGACATTT TTCTTTACAT AACCATAATT CTATTATTAC
17501 ATCTTATCTC TGTGTTGTCT AACACCCAGT CCATATTCCA GTTTCTCTGA
17551 TTGTCTAAAA ATGTCACCTT GTATTTGGTT AAGTTTCTTA AGTCTCTTTT
17601 AATCTTTAAG CATAATGTAT TTCTTTTTTT TAAGTCCTCT ACATAAATAAT
17651 GACATATTTT ACAGATTTGT TTAATGCCTC TGTAGGTTAG TGATTTACAG
17701 CTAGGGATGA GCTCAGGTAG TGGGATTATT TGATTTGAGA GAGGAAATAC
17751 AGCTATTATA AAGATTTGGA AGTAAATCCA TAACTGAAAG CCAATGACAG
17801 ATCTTTTTTC CCTTCTAGGT AAATGACTTG CTCTGGAATG CAGATTCCTC
17851 TGTGCTTGCA GTCTGGCTGG AAGACCTTCA GAGAGAAGAA AGCTCCATTC
17901 CGAAAACCTG TGGTAAGACA GCTGTAGTAC CCCAGCCTTC TGCCCCATAA
17951 AACGTAGTTG AAAGTAGACA GGTATGGGAT TTCCTTCATC CCTTCTACTT
18001 AGTCCCTTAG TAGAATCAAA GATGCTGAAG TGGGTAGGTG GAAATGGGGG

594221_1

18051 TGGTTAGGTT TTGATTGATT GTGGATTTCA GTCATGTATT GGTTGGGGTT
18101 CTCTAGAGAA ACAAATAATA CATATATATA ATTCGTCCCT CAGTATTCTC
18151 GGGGGATTAG TTCTAGGATT GCCCATGGAC GCCAAAATCC ACACATGGTC
18201 AAGTCCTGCA GTCAACCCTG CAGAACACTC AGATATGAAA AGTCAGCCTT
18251 TTGTATACTT GGGTTTTGCA TTCCTCAAGT ACCATATTTT TGATGTGCGT
18301 TTGGTTGCGG GTATAGAATC CACAATATGA AGGGCCGACT GTATTCATTG
18351 AAAAAAATAC GAATATAAAT GGACCTGTGT AGTTCAAGCC TGTGTTGTTC
18401 AAGGGTCAGC TGTACTTACA TAGAGAGACG GTGAGAGAGG GAATAGGGTG
18451 GGGCGGGAGG GAGAGAGAGT AATAGAGTGT GGATAGATTT ACTTTAAAAG
18501 ATTAGCTAAT GTAGGGGATG GCAAGTTTGA AATTTGTGGG GGCAGGTTGG
18551 CAGGCTGGAA ATTCAGGTAA GAATTGATGT TGCTGTCTTG AGTATGAAAT
18601 CTGTAGGGCA GGCTGGAAAC TTAGGGAGGA TTTCTGTTAC AGCCTTAAGG
18651 CAGAATTTCT TCTTTTCTGC GAAGCCTCAG TTTTTGCTTT TAAGGTCTTC
18701 AGCTGAATGA ATGGGACCTT CCCACATTAT GGGGAATAAT CTGCTTTCCT
18751 TATAGTCAGC CGATTATAAA TATTAATCAC ATCTACAGAA TACCTTCACA
18801 GCAACATCTG GAGTTTAGCA GATAGCTGGG TGCCATAGCC TAGCCAACTT
18851 GACACAATAA AATTAACTGT TGTAAGTCAT CACGTGCTTT CCCTAGTGCA
18901 TGGTATTACC ACAGAAAAAA CACTAACCAA AGGAATTCTG TGGACGTGAA
18951 AGAAGATTTA GATTAAGCGT AAAAGTAAGA ATATTTTTAT AGCTTTTAAA
19001 ATGTATAAGT GTGTGGTTTT AAGTATTAAA TAATACTTGA AAATGTTAGA
19051 AAATAAGATG AGAAAAAAAT CTCATAGTTC TACCACTTCG TAATAATCAC
19101 TATTCAAATT TTCTTGTCTT CTAGGTTTTT CATGTATATA TCTCAGTATA
19151 GCTATCATCT TGTTTTTGTT AAAAGTGTAG TAGGTATGGG CCAGGTGCGG
19201 TGGCTCATGC ACTTTGGGGG CCCAGCACTT TGGGAGGCCG AGGCGGGCGG
19251 ATCACGAGGT CAGGAGATCG AGACCGTCCT GGCTAACACG GTGTAACCCC
19301 ATCTCTACTA AAAATACAAA AAATTAGCTG GGCGTGGTGG CAGGCGCCTG
19351 TAGTCCCAGC TACTCAGGAG GCTGAGGCAG GAGAATGGTG TGAACCTGGA
19401 GGAGGCGGAG CTTGCAGTGA ATGGAGATCG TGCCACTGCA CTCCAGCCTT

19451 GGCGACAGAG TGAGACTGTC TCAAAACAAA ACAAAAAAAA GTGTAGGTGT
19501 GATACATCTG CATCATTTTA AATTGCTGTA TAATACTCGT TTATTCTCGT
19551 TCATTAAATC TCATGCTGTT AGACATTTAC AGTTTTGTCA TTTCTCATTA
19601 TTGTAAACAG CAATGCATGG TACATTTTTG TTCATAAATC TTTTTACTTG
19651 ATTATTTTCT AAGTAGCTTT CAAACTCTTT AATCAGTAGA ACCCCCCCCC
19701 TTTTTTTTTT TTTTTGGAGA CGGAGTCTCT CTCTTTCCCC CAGGCTGGAG
19751 TGCAGTGGCC CGATCTCGGT CACTGCAAGC TCTGCCTCCC GGGTTCACTC
19801 CATTTTCCTG CCTCAGCTTC CCGAGTAGCT GGGTCTACAG GCGCCCGCCA
19851 CCAAGCCTGG CTAATTTTTT GTATTTTTGG TAGAGGCAGG GTTTCACCGC
19901 GTTAGCCAGG ATGGTCTCGA TCTCCATCTC GTGATCTGCC CGTCTCGGCC
19951 TCCCAAAGTG CTGGGATTAC AGGCGTGAGC CACCGTGCCC GGCCTCAGTA
20001 GAACCCTTTT AACTGCAATG TTAAGAAACT CATTATTCAT TCAACACAAT
20051 AGTTCTTAAC CCTGGCCACA CCTTTAGAAA AAAAATGATA TTCAGGCTTC
20101 ATCTAAGAGT TCAGTTCAGT GTGTTGGAAT GGAGATTATA CGTAAGTATT
20151 TAATTAAAAA CCAAAAGCCC CCAAGTGATT TTAAACAGCC GCAGTTGAGA
20201 ACCACCGATT AACCAGTGTG TCAAGGGATG GCACTGTGAT ATGCTGAGCA
20251 TAAAAATATT GCACAGGATG AAACCCTGTC TCTACTAAAA ATGCAAAAAT
20301 TAGTCCGGCG TGGTGGTGCG CGCCTGTAGT CCTAGCTACT CGGGAGGCTG
20351 AGACAAGGGA ATCGCTTGAA CTGGGAGGCA GAGGTTGCCG TGAGCCGAGA
20401 TTGAGCCACT GCACTCCAGC ATGGGTGACA GAGTGAGACT CCATCTCAAA
20451 AACATGTATA TATATATATA CACACACACA CACATTGCAC AAGAACAGCC
20501 ACAACATCTG TGCTCACAGA ACATCAGCAT GTGGTCTAAC TTCAAAGTGT
20551 TGTAATAATG CGGTTTGAGA CTAGGTTATG TTTGCTGTGA TCACTAAGTT
20601 AAGCATTAGT GAGCAAGGAG ATTGAGAAAA TCCTTAATAT AAATAATATT
20651 TCTTAATATA ACTATAATTC CTAATATAAC TAAGGTCTTA ATTTATATGT
20701 CATCTGTTTA GTAAAGGTTG GTTTTGGCAT GATTAAGTCT TGCTTGCTTA
20751 ATAGATGTTG GAAGGATAAT TTCATGCTTA TCTTCTTTGG ACAGCTGAAT
20801 CAGGATTAAT ACCCAGATAG CCTTGAACAT AAGTGCTTGC AAAGCACCTG

20851 AAAGAAAATA AGCATCTTAA GCCCAATACA ACACAATGAT GCTAGTCTAG
20901 ATCTTGGATT AAGTGTTTTA ATACTTTTAC TCTAATTGCC AAGTTATCTT
20951 CTTCCTAAAT CTTCATGAGA AAACCCACTA AAAGAATGCT TTTTCCTGGT
21001 AGCCTTCCAT TGTGATCATA AAGTTTGGAA GTAAAGTTGA AAATAAACAT
21051 GTGGGCCAGG CACGGTGGCT CAGGCCTGTA ATCTCAGCAC TTTGGGAGGC
21101 CGAGGCAGGC GGATCACAAG GTCAGGAGAT CAAGACCATC CTGGCTAACA
21151 CGGTGAAACC ATGTTTCTAC TAAAAATACA AAAAAAAAAA ATTAGCCGGG
21201 TGTGGTGGTG GGCGCCTGTA GTCCTAGCTA CTCGAGAGGC TGAGGCAGGA
21251 GAATGGCATG AACCCGGGAG ATGGAGCTTG CAGTGAGCCG AGATTGCGCC
21301 ACTGCACTCC AGCCTGGCCG GCAGAGCGAG ACTCTGTGTC AATAAAAAAA
21351 AAAAAAAAAC GAAAATAAAC ATATGAATAA AAGTTAAAAA TAGAAAAAAA
21401 ACAAGAAAAT AAACATATAT TTCTGACCTT ATTGATTCTT GATATTTTAT
21451 CTGCATGGAA AGCTATTTTT TGGCAGTTAT TATTGTTCTT ATTTTAGAGA
21501 CGAGGCTGAG CAGGAAGGGT CCTTTGAAAA AGAAAAGATT GCCCTTGAAC
21551 CCCTCTGGCA AGTGGGATGA AGTCTGCTTC CCAGCCTCTA ACGGCCTTCT
21601 TTTCATTTTC CCTTGCAGTT CAGCTCTGGA CTGTTGGAAA CTATCACTGG
21651 TATCTCAAGC AAAGTTTATC CTTCAGCACC TGTGGGAAGA GCAAGATTGT
21701 GTCTCTGATG TGGGACCCTG TGACCCCATA CCGGCTGCAT GTTCTCTGTC
21751 AGGGCTGGCA TTACCTCGCC TATGATTGGC ACTGGACGAC TGACCGGAGC
21801 GTGGGAGATA ATTCAAGTGA CTTGTCCAAT GTGGCTGTCA TTGATGGAAG
21851 TAAGCTCCTG GGAAGTGTGT CCATGAGCCT GCAAGGGGTC CTGAGCCTAG
21901 GGCCTGCAGA TGTGGTGGTT TGACTGGAAC AGTGGGGAAT CTTTATTTGT
21951 TTTGGCTGTT TGGGTTACTT GTTTTTTTAT TGAATGGGAT ATAAGGTGGG
22001 GTATGTTCTC TCCTGAGAAC CATTGTCCCC CCTCCCCCAC CAGTTTCCTG
22051 TTATACTGCA TCTGTGGCCT TCACACGTTT ACTTGCCTGG CCTTTGAAGA
22101 CACTGAAAAC TTTGACTCTA GGTAGAGAGG ATGACAACAG TACAGTCTTG
22151 TGGGATTGGG TGTGTTAGCT TTATCTGTTT GCCCTGACAC AGATTTATAA
22201 TTGACCCTTA TACCACCCCA CTTGTGTTGC TTTGTTTCCT GATACAAATG

22251 CTTGCTGATA TATACCTCTC CAGTATGTTC AGTTCATGCA TAAACGTTTG
22301 CCTAATATGA AGATTAGGTT TATATTTTAT AATGAGGTAG AAGGTTTTTT
22351 TAGGGGGTGG GGTGGGAAGG GCAAGACTGA AGAGTGAAGT AGTCACCTTA
22401 ATGAATAGTT TCATTGCTGA TATGAAAGGG AGCACTGGCT TCTAAGATTG
22451 TAATGTGAGG TGGATATTAA TTCATATTCT GTGTAATATT CTACATAATA
22501 CTGATTTTAT AGTCATGTAT TCTATATAGA GAACTTAATC AGATCTGCGT
22551 TATTACCAAA TCCACACATA GGAAAGTGCT TTAAGGATTT TGAAAGTATT
22601 AATTCCCTTG GTTTAGTGTG GCTTGGTTGC AGGCCCAGGT TTAAAGCTAG
22651 AGGTCTGACC TCTTGGCCTT TTTGCCTTAG TCCCTGGCAC CTGAAACTCC
22701 AGGTACTGAG ATGGACTCCC CTAGGCCTAG AGGTGACAAT AGCCAATTAT
22751 GGACAGAACC CATGACATTT CCCCATCCCA CACTGTTTTT AGACTTGTTC
22801 CTGAGAAAAA CATTGAAAGT TATTTTTTTG TGAATTGCCA TTATTGTTTA
22851 GATATACTGT GATGTTCAGA TGGCTTATCT TACAAATTGA ATATCCCTAG
22901 GTCTAATCCT CTTCTTTCTT TTTCACTGCA GACAGGGTGT TGGTGACAGT
22951 CTTCCGGCAG ACTGTGGTTC CGCCTCCCAT GTGCACCTAC CAACTGCTGT
23001 TCCCACACCC TGTGAATCAA GTCACATTCT TAGCACACCC TCAAAAGAGT
23051 AATGACCTTG CTGTTCTAGA TGCCAGTAAC CAGATTTCTG TTTATAAATG
23101 TGGTATGTTA TAAAACTTTT GCCAAGATGT TCTGAATCAA GTCCCTTCTA
23151 CTCCTACATA AAAGCAAATT ATAGTTTGGT GTTGCCATAG GTCTAGTGTT
23201 TCTCAAAATT TTTAAGTCTG CAGTTGATAT CATTATCATT ATGATATTTA
23251 ATTGCCTTGG GTTTTTGTTT TTTTTTTTTT TAATCCTATA CTGGTTTGTA
23301 CGAGCCATTC CTTTTCCCTT ACTGACTTGA AGAGTCAGTT ATTTAAGAAT
23351 AACATTGGAC TCTGGAAATA ACATAGTATG TTATACATTG TTAACATGTT
23401 TTACTCTTTT CATAGCCTTT ACACATATTT TCAGTTGATC TCATCCCTCC
23451 TAGGAGCTGT GTCAGAGATG GGGTTTTCCT CTTTTGTAGA TGAGGGAACA
23501 CAGTGTCAGA GGTTTTGTAA TTTGTTTGAA CAAGAATGGA CAAGGACCTC
23551 AACACAGGTG TTCTAGCTCC TAATCCACTT GTCCTGCCAC AGCCCCATTG
23601 CTGTCAGTTC TTCATTACTT TCCTGATGTG CTGGAGAATC TGAAATTTGT

23651 TTTTACTTGT GAGTTCTGTG GTTATGTCAT AAATTCTGCT GGCATATGGC
23701 AGTGTTAGCC TTGTTTTCAA ATATCTTTTG AATTCTCAGA AAAAGCCTAG
23751 ATAGTTGCCA AGAGAGAATA ATCAAAATTA ATTAATTTAA ATGGGAAGTC
23801 CTTACTTTCA TATCAGCTTT TCTGTTAAGT CAGCAGCCCA CTGTGTACAT
23851 GGATCCTATC TGGATGTATC ACCAGTTTCT CTGATTATAG TTTCAGTGTG
23901 TAAAATGCTG TTACAGTCCT CCTTAAACTT TTCAAAATAG CTTTAAAAAA
23951 AAGTGCAAAT ATGTTCATTG TCAAGGCAAA AAGAATCAGA TGTAAGCTTT
24001 TGTGGGACTT AACTGTATGA TGCTAATGAG TTTATATGTC ACTTTATGAT
24051 GTATGGTATG TTTTGTTCTG CATTCACTTA AAAAATAGCT TTATATCATT
24101 CATCTATTTA AAGTGTACAA TTCAATGGTT TATATGTGTG TGTATGAATA
24151 TATATACATA TGTATATGTA TATATATGTA TATTCACAGA GTTGTACAGC
24201 CATCACCACG ATCAATTTTA GGACGTTTTT ATCTCCTCAG AATGAAACCC
24251 TGTACCACCC TGCATTCATT TTACTTGAGA GAAAACTCCC TGTGATGAGA
24301 TAGGACAGGT TGAGAGCTCC ACTTTTGAAA GATTGTTCGG CATCAATATG
24351 TGGGGTTGGC CATAGGTCAG GGGCACCTGG AGGCAGAGAT TCTAGTTAGG
24401 AGAAGCTGTT GTCAAGTGTC CAGGCAGGAG CTAGCAAGAG CTTGAGCCAG
24451 AGCAGTGTTC ATAGAAATGG AAAGAAGAGA AAGATCATAA CAAATCCATG
24501 AAGTAAAAAC CCTGAGAAGT TAAAGAACCC ACTGGGGAGA GTTTGGATAT
24551 AAGAGAATCT GGAAAAAGAG ATCTTGGACT GGAACAGGTC AGGGCTCCGT
24601 GCCCAAGTGG AAGGGAAATT AAGAACTTGG AGTCAAGTGG TAGACATTTG
24651 AGTGGTGTGG AGACAAGTTC GTTGCCAAAG TTTTCAAAGA TGGTGTTTGA
24701 TGCATCCTGA GTATCACTCC TTTTTCCCCC TCATTGCTTC TTGATTGTTT
24751 ATTATATGCC AGGCTTTTTT CTAGTACTTG GCTTGTTGTA CTAGAAAACT
24801 AGTTGTACTT TGTCTACAAC TTGTTGTTCT AGGTGTAGAC AAAAGATATC
24851 AATTAAATAT GATCTATCAG ATGGCAAGTG CTGTGGAGAA AAATTAAGCA
24901 AAATAAGGGG TAGGGAGAGC TTAAGGATAA GGGTTTACAG GGGGAAGGTG
24951 TCTTTCCTAT TTAGTGTGAT CCCAAAGGCC TCTCTGTGAA GGTGACATTG
25001 AAGCAGAGAC CTGGTGAGAA TCACAGTGGG AGCCACGCAG ACATCTGGGG

25051 TAAGAGCGTC CCAAGCATTC TATGCTTGAA GGCAAAGAAG AAAAAAGAAA
25101 GAGCGTTCCA AGCAGAGTAA AAAGCAACCA CCGAAGTGCC TGTTGTGTTT
25151 AGGAAATAGC CAGGAGGCCA GGGTGGCTGC AGCAGAGCAA AGGAGGGGAA
25201 GGTGGTGGGT GAGTTCAGAG TGGTGATGGG AATCTGCTCT TGTAGGGCCT
25251 TGCGGCTTTT ACTCCGAGTG AGATAGGAGC CACCAGAGGG CTTAGAACAG
25301 AGGAGTGCAG TGTTCTGGCT GAATTTTTTA AAGGCTTGCA TTGGCTGCTG
25351 TGCAGTGAAT AAACTGGATG AAGAATAGAA AGAAAATGTC TTTTAAGCAG
25401 GTGCTTAGGA CTTTGGAGAA TTTGAGGATA TTGAGAGGTG GTTGAAGACA
25451 GTGGAGGAAA TTGTCCACAG CACTGGGCTG AGAGGGTAGC CCCTTCACCT
25501 GGTCTTGCTG AGATGTGGCC TTTGTCAGGG AAGATTATGA CTGATGTGTT
25551 CTTAAGAGGA AAGCAGAGAT TTTAAGGAGG TTGAGATGTG ATTATTTTCT
25601 AGATTGCTGT TTGCCTTCTA GAACTCATTA ATTGCAGACA CCATCCCCTT
25651 AGTATTAGGT GAAATCTTAT AATTTACGAT GATAATATTT GCATTTTTGT
25701 TTTCCAGGTG ATTGTCCAAG TGCTGACCCT ACAGTGAAAC TGGGAGCTGT
25751 GGGTGGAAGT GGATTTAAAG TTTGCCTTAG AACTCCTCAT TTGGAAAAGA
25801 GATACAAGTA GGTTCTTAAT TATCTTGGGC TTCTGGGAAC AGAATCAGCC
25851 AGCATGCAGT CCTAAATTCA GCCATCTGAT AACAGTTCTA TGCCTGTTGC
25901 TGAGTGGAAC AAGAAATAAA GACAACACCC AGGCCCTGAC TTTCGGATCT
25951 GATTGGAGAA GCCAGTCATG TAGTTTGTCT GAATGCCATA TAATTTGATA
26001 GGTAGCAGGA GAGCATGAGT TGTAAGCCAG CCTAGGACCT ACTCCCAATA
26051 GCGCTTGGTT CTCCAGGAAA AATCATGTGG GAAAGATGGA GATGACAATG
26101 ATAAGGCGGA GCTGCATTCT CTTACATAAA TGGGGATGTA TGGGTTGTTA
26151 ACATGGATGA CCTAATGCAG CCTCTGTCTT TGCTCCATCC CAGAATCTAG
26201 AACTTCTGGG TGCTGTGCTT TGAGGCTCCT GGGATGGAAA TCAGAATGCA
26251 TTCTTCCATT GAAACAGTAT TGTAAACAAT TGGATGTTAT TGAATACCTC
26301 AGGTACACTA TAGGCATTTG CAAAATGACC TAGAAACCAA ATTATAATGC
26351 CACATCTGTG AGAGAACTTT TTTAAAAAGT ACCACTTATT GAGTACTTAC
26401 AGATTAAAAA AACAAAGTGT AGAGGTTAGG TAACTTACCC AAGGTCATGG

26451 ACCTGGTAAC TAGAGAATTT AGGGTTTGAT TCTATTCTGT TTGATAAGTC
26501 CATGTTCTTC ATTACTAAAC TACTCTGCCT CCAGGGAACA TTTATTGTTA
26551 GATTAATAGA AATAATTAAC TGAGTACAAC AAATAGCAGA ATTTAATAAA
26601 TAATGTTTCT TAAATATATG TGATATATTT AATAAATACA GCAGAAGTGT
26651 TCAACCTCTG TATGATTTTG AGGCTGCCTG TATAATGCTT AGTAGTTTTT
26701 AAAGAGCATT TACATGCATT ATTTCACTTC ATAGACTTGA AACCACTAGA
26751 GTAGAGATAG AGGACAAATT AGAAAGTATG AGGCAGTTTA GAATATAGTT
26801 TCATTTAAAA AAAATTGATG GGGATAATGC CAATTCGTCT GAGATTTCAC
26851 AGAAGACATG AGTACTCATC GTGATCTTGG GGAAGGGATA GGTTTGGGGT
26901 TGGCAAAGAA TTGGGAACAT TGGGTCTGGT GGGGAAGAAA GTGTCAGTGA
26951 AAACCAGAGG TGGGACTGAT CCTCCATGGG ATACTCTATG TGAATGCAAT
27001 GGAGAGCCTG AGTCCGGGGA GAGATGTTTG AGGAGGAAGA TCAGGCTAGT
27051 GACCAACTTC TTCAGTGGGA GCTGCGGATT TGCCACCTGA TCTAAAAGGC
27101 AGGAAGTAGC CATTGTCGGT TCCTACGTGA GGTGACAAGA ACAGTGCGCT
27151 GGTCAGGTGT ATAAATGCTA CCAAAGAATG CATTAGAGAC ATGGAGACCA
27201 TCTCTCAAGC TAGTCAGTCA GTTTAATGTG AGGTGCTTAG GAAAGGACCC
27251 ATTCTACTGC AAGTGACATA CCTGCCAGAG CCTGGTTTGA ATGCTGGTAA
27301 GTCATGGCAG TGGAAAAGCT CTGGGGTTCA TTAGTGTAGG GACTAGGGCT
27351 GGTAATTTTC TTGTGTAGTC AGTTTCCTCA AGTGTTCTCT TCAAATTTAA
27401 AGATTTCAGG GTATGAGAAA TTTAGGGAAA ATATAAAAAC GTATTCTTAA
27451 GCCAGACAAA GATTAATTTT AGATTTTGTA GTATTTGGTA GTATCTCAGG
27501 TTTTGTCCCT CCAAATAATT AGGAGTGGAC TGTATACAAG ATGCTTCAGT
27551 CTTCCTTCAT CCAGGAACGT CTCAGTGGTT TTTAAGTTTT ATTCATGTCT
27601 TGGATATTCT TCAATATTTA CAATAGAATC CAGTTTGAGA ATAATGAAGA
27651 TCAAGATGTA AACCCGCTGA AACTAGGCCT TCTCACTTGG ATTGAAGAAG
27701 ACGTCTTCCT GGCTGTAAGC CACAGTGAGT TCAGCCCCCG GTCTGTCATT
27751 CACCATTTGA CTGCAGCTTC TTCTGAGATG GATGAAGAGC ATGGACAGCT
27801 CAATGTCAGG TATTGCAGTT TTTCCCTGTA CTCCACATGT TAAGCAAATG

27851 GAGTTAGGTT TTTGTCTTTT ATGAGCATAC AACTTTTGAC TTCTATTGAT
27901 CAAGGTTGAG GAGCAGTAGC TTTCTTGTTA GACACACTTA ACAAGAAGGT
27951 TAAGTCTAGT TATGAGCCAT GTCAAAATAA CAGACCAAAA ATATATCAAA
28001 AAGTGGTGAA AAATAGGATA AATATTAGTA GATGAAGCAA CTTTTTAAAG
28051 ATATGTTAAA TATTTTAATT TAGCATCTAC CCACATTTTT CCAGCGTGAT
28101 TGTTATATGT TATAATTGAT TTTAATAACT GTCAAGCATA ATTAGAGTGG
28151 CTAATTCTCA TGGGCTAATG TGATGGGAAG AAATTTTGTA TAAATGCAGT
28201 CATGCGCATA TATGTGTGTG TGTGTGTGTG TGTGTGTGTG TATACATACC
28251 TTTTCTATGT TTAGATACAC AAATACTTGA CATGGTATTA CAATTGCCTG
28301 TAGTATTCTG TAAAGTAACA TGCTGTCCAG GTTTGTAGCC TGGTAGCAAT
28351 AGGCCATACC CCATAGGCTA GGGGTGTAGT AGGCTACACC ACCTAGGTTT
28401 GTGTAAGTAC TCTATGATGT TTGCACAATG ATGAAATCAC CTAACAACAC
28451 ATTTCTCAGA CGTATCCCCA TCGTTAAATG ATGCATAATT GCACATATAT
28501 GCTTTGTTTT GATGTGGTGA CTTCAAAATG CTTCTTCCAG CCTCCTCTTC
28551 TATATATCCT ATTTTGTACC TGACTACATT TACCATTAGA AAGTCTCTAT
28601 TCTTCTTTGC TGAAATTTCA CTGTTCTCTG GGCCTGAGTT TTGTTTTGAT
28651 TCCTGACTAT ATCTTCATTA TGTAACAGGT TTCAGTTAAT GAATGCTCTT
28701 CTGTGTAATG TAAGCCCTGT TGTATAGTTG ATAGCATTTT CTAGCCAGTT
28751 CCCAGAACTC CTTGTTTCCA GTGTCAATAC TTGGCACCTT TGTCCACTGA
28801 CACTAATCCC CAGATTAATT TGTAATTAAA GCCCTACTGG TGAGATTTCT
28851 GAGAAACGTT GTTGCAAAAT TAGGAACCTT TCCTTTATAT ATATACATTA
28901 CATAAATTTA TAGACATAAA ACATTTTAAT GCAGTCATTT GCTGCTACTC
28951 TTTGACTCAT AGTCTTTCGT GATATTTTGA AAAAGCCTTT TGTTAACATG
29001 TCTAAATGCA GAATATGTTC TAGAAATATG TAGCACTTAA AGTAAGCCAT
29051 TAGATTACCT TTTGAAAAGC GGAGCAATTT ACTAAGTTTC TACTTCTTCA
29101 GATTTGAAAT TCTTCATCAT TAGCTTGTAG AGGCAAAAGC TTGATGCAGT
29151 CATCTCATTT GCTGTAAAGG AAATGAGAAG TCATTTACAG TATATTTCTA
29201 CTGCTTTGAC TTTTATTTCT CAAAAAGACT GTTTTGTTCA TATAAAATAT

29251 TAATGCTTTT GAGGACTACA AAGTCCCTCG ATTTAGTTTA CATTTACTTT
29301 AGCTTATACT TTGTAAAAAA TACTCTTCTA AATGCTTTGT CTGTTTTAGC
29351 TTACTTATTT CTCATAATAC CTCTGTAAAG TATATGCCAT TTGCACCATC
29401 ATTTTACAGA TGAGACAACT AAGACATGGA GCAGTTAGGT AACTTGCCTG
29451 AGATCATGCA GGTGGAGCCA GGATCAAATC CCAGCGAGTC TAGCTCCAGA
29501 GTTTGTTCTC TTCTTGACAG ATAATTTATC CTCACAAAAT TTGAAGCATT
29551 TGTAGAGGAA TTCCCTATTG TTATAATGTT TAGTTTTTTT GTAGATTGGT
29601 TAAAAACTTT GAATTAAATG TTAGCATTAA CATCATTTGC TTTTATCACT
29651 ACTTCTTTGT CTCTTTTTTC TTTTTTTAAT CACTACCTCT TCCTCCTCTT
29701 TTGAGAAATT CTGCTTCCGT GGCTATGGTC CAAGCTACTT GAGAAGGTGA
29751 GGTGGGAGGA TCACTTGAGC CTAGGAGGTT GAGATTGCGG TGAGCTGTGA
29801 TTGTGTCAAC TGCATTTCAA CCTGGGCAAC AGAGCAAGAC ACTGTCCAAA
29851 AAAAAAAAAA AAAATAGTGA AATTTTACTT CGCTCCATTG ACTCAGGGAA
29901 AAAATGTAAT GGTGATAACA AATTCCCTTC ATCTCATTAG TGAAAATCCA
29951 CAATTTTCCA TCAATCGATA TGATAGTGAT AGAGATATTG AGTGTGCTCA
30001 TTTTCCTACA GACCAGCTGC TTTAACTATT TTAAGCAGAC AGAAATGATA
30051 TTGGTACCAT CCATGTCTAA TGAAGGCAAT ACTTTGTAAT AAGTTGCAGT
30101 AAGTTGTGGC CAGAAGAGGA ATGATGACTT CACAGTGTAA ACAACTACCT
30151 TATTGGGTTT GTGGAAAATG GTGTCATGTA GCAGATGTGG CTTTATCTGG
30201 GCTTTGGTTT GGAGTAGTTT TATCTATTCA TCTAACCGTC TGTCTCTAAG
30251 TGTATAAGTG TGTGTGTGTG TGTGTGTATA GTATTGGGTG TGTATATATG
30301 TATTTTGTCT ACATTGTATT GAAGTAGGTA GTGCAGCATC AAAAGGAAAT
30351 TGTTGATTTT CAAAATCAGT GAAATGTCAC TATTTTTGAG AAAAATGGTC
30401 TGTTTACACT CCCTTCTCCT TTTTTTTGTC AGTTCATCTG CAGCGGTGGA
30451 TGGGGTCATA ATCAGTCTAT GTTGCAATTC CAAGACCAAG TCAGTAGTAT
30501 TACAGCTGGC TGATGGCCAG ATATTTAAGT ACCTTTGGGG TGAGTATCAA
30551 GGTGTTAGGA AAGCATGTTA TGACTTACAT AGATGCTTAG TTCTTAAGAA
30601 CATGTACTTG TATCTTGTCA GTTCAATATT GATTGTCAGG TCTTTTAACT

30651 ACCCTGGAAA ACCCTAAGCT TTAGAGTGGA ATTGGCAAGT GTATTCTACT
30701 CCTGTTTCCT CTTTTAATGA ACTAACGTAC TCTTAAAAAA GTGATTGATG
30751 ACTATCGCAG GGACAAAAAA CGAAACACCG CATGTTCTCA CTCATAGGTG
30801 GGAACTGAAC AGTGAGAACA CTTGGACACA GGAAGGAGAA CATCACACAC
30851 TTGGGCCTGT CGTGGGGTGG GGGAGGGTGG AGGGATAGCA TTAGGAGATA
30901 TACCTAATGT AAATGACAAG TTAATGGGTG CAGCACACCA ACATGGCACA
30951 TGTATACATA TGTAACAAAC CTGCACATTG TGCACATGTA CCCTAGAACT
31001 TAAAGTATAA TAAAAATATA TATATAAATA AATAATGCCA GCATTAGAGA
31051 AAAAAAGTGA TTGAAATTGC ATGTTAAGTG TTTTAGCAAA TGTTGATGTT
31101 GATGGTTTTT TGCAAAGAGC GCATCAGCTA TTTGTGAACT AGATCTGTGA
31151 ATCTTGCAGA GTCACCTTCT CTGGCTATTA AACCATGGAA GAACTCTGGT
31201 GGATTTCCTG TTCGGTTTCC TTATCCATGC ACCCAGACCG AATTGGCCAT
31251 GATTGGAGAA GAGGTAGGTG AACACGGAGC AGGAAATTTA CTTAAAGTAG
31301 TTACCCAGGG ACTGATGGCA TTAAGTAGAA AGAGCGTGGG CTTTGGAGGT
31351 GGACTTGGGT CTCCACTAAA TGCCTAGACA ATAGTGGGAA ATGATCTCAC
31401 TTTCATAAGC CACACCTTAT TCATCTATAA AATGGGAAAA TCAGTATCTG
31451 TCTATCAGGG TTCAGAAGAC TAAATGAGAT AATATATGTG ATTAGCAACC
31501 TTTTATCCCT AGTTGTACAA ATCATTCAAA GTTAATTTTA TTTAGTAGGG
31551 GAAACAGAAA TGTGATCTTG AGAATAGTTT TAGTAGATTT TTATTCAACA
31601 CATACTAGAA TGCCTATAAT TGTGGTGGAT GGTAGAATGC AGTGGCTGGA
31651 AAACAAAACC GCTTGACTAA TTCCTGCTCT TCTGGAACTT GTGATCTATT
31701 AATTTCAATG TAATGATTCC CTTTGTTGGG AGTGTGATGG AAATGGACAG
31751 AGTATACTGG TAGAGAATAC TGAGATGTTT GAGGGGTAAT TTGAGGATGG
31801 TGGCTATGAG AATGGGAGTC CTGCATCTGG TGGTCCAGGA AGGCCTCTCG
31851 GAGGCAGTGA TGTGTGTGCT GAGATGTGAA GAAAAAGAAG GCTCTGTCTC
31901 CAGGCAGAAG GAACAACAAA CTCCTTGAGC TTAGCAAGAG CTCATCTTAT
31951 TCAAGGGACT GGATGGAAGT ATTGTGGCTG GAGCTCAGTG ACAGTCATAG
32001 GAGGGAATTT GGGTTCTTTA ATTGAACAAA GATTAGAAAC TTCTTGTGAT

32051 TTTTAATAAC AGAGTAATGT GTTCTGCTTC ATGGTTTGGA CAGTGATTCT

32101 GGCTGCCCAG AAGAGACTTG ATTGGAGAGT GACGAGACTG GAATATGGGA

32151 TCAACACCGG TTGAGTGGAG TTAGTGAGGG GAAAAAGGAG ATGGGTTTGA

32201 GATATGTGTA GGAGATGGAG ATGTCAGGGC TCACTGATGG ATTGGATGGC

32251 TTCACATTCC GTTTTGCACT GGACCAGCCA CGTCTTAGGT ATCTATCTTT

32301 AGTCCTGATT ACAGGAACTT AGGTGTGAAA TCATAGGGTG GTAGAACTAT

32351 GTGATAGAAA AGGTAGGTTT AACTGATTTG AGATAGAATT GCTTGTGATT

32401 TCAGTTTTAT TTCTTTGCAG GAATGTGTCC TTGGTCTGAC TGACAGGTGT

32451 CGCTTTTTCA TCAATGACAT TGAGGTATCA AGGCTTGGTT TGGTGTTGGA

32501 TCCTTTTCAC AGTGTTAGCT CCGAGTAATC TAGCTAGCTT TCACCCATGC

32551 CTCTCTGGCC TTCTCTTGCA GGTTGCGTCA AATATCACGT CATTTGCAGT

32601 ATATGATGAG TTTTTATTGT TGACAACCCA TTCCCATACC TGCCAGTGTT

32651 TTTGCCTGAG GGATGCTTCA TTTAAAAGTA AGTTTTCAAT GTATAAAACA

32701 GAAATGGTCC CTTCTCCAAT GTCTTTTGGA GTCTTGATGA CTTTTTGAAT

32751 TCTTCATTTA TTTTGGCTTT TTATCAAGGA GTCCTAGGCT GGAGAAAATC

32801 TTTAGAGTTA TTTTACTTAG ACCCTAATCT CAACATAATA TCTCAGTTAA

32851 ATCATTCTGC ACTTTAGTAA AGACATCCAA GGAAGGGAGT TCCTTCCTTA

32901 AGCAGCACAT TCTAAAGTTA AAAACTTTTC AGGAAATTTT ATTATGTAAC

32951 TGATCTAATA TTTTATTTGG AATTACTATG TAGATCCCCA ATGTTTTACC

33001 TTCTGTGTAG TCTTTTCCCA CTGTGCCCAC CCTCCACTGT ACATCTGCGC

33051 TCCATCTAGT GGTTTGTAGG ATATTGGCTG CATTTTGTCT TCTGTTCCAT

33101 GCCCTATCTA TCTCTGTGTG TGTGGCGTGT ATGTGTGTGT GGCGTGTATG

33151 TGTGTGTGGC GTGTATGTGT GTGTGGCGTG TATGTGTGTG TGGCGTGTAT

33201 GTGTGTGTGG CGTGTATGTG TGTGTGGCGT GTATGTGTGT GTGGCGTGTA

33251 TGTGTGTGTG TGTTCCTTAT TCTAAAAAGC CAACTTATTT TCTTTGCTTC

33301 CAACTTGGAA ATAGGGAATC TTTCTTTCAT TGATATGATT ATAGTACACT

33351 GATAATGCTA AGAAATAGAG AAGTTGCCCC AATTCTTAAC TGTGTTTCTC

33401 CACATCATTT GAGAAGCTGT GTATGTGAAT GTGCATGAGG GCTCTGTAAG

33451 AGAGAGGGCA AGTTCCAGGG ATGAGCGTGT TCATCAGCAG GGCTGATAGT
33501 CTTGAGGTTC AGTGGGAGAG CTAAGGCACA TGGTTGTTAT TTGTTCTCTT
33551 CTATTTCACA TAATGTGTGC GGTTTCAATT GCAGTTAATG GAGAGTGGCT
33601 TGTTGTGATA ATTAAGGCTT ATTAGTTAAT GGTGTGTTTA GCATTACAGG
33651 CCGGCCTGAG CAGCAATCAT GTGTCCCATG GGGAAGTTCT GCGGAAAGTG
33701 GAGAGGGGTT CACGGATTGT CACTGTTGTG CCCCAGGACA CAAAGCTTGT
33751 ATTACAGGTA AGCTGGTTTT TCAGACAAGA TAGATAGTCT GATTGTCATT
33801 CAGCCAAGTA CCAAGCATAA TTCTTGCAGG TTGTATTTTA GGCTTTCTTA
33851 TTCTTTGTAT CGTTTATTGT AAACCTTTCC TTGATAGTTT TCTGTTAGCT
33901 TTATTCAAAG GAGTGTTGAT ACAGGCTGTG ACCATAAGGC TCAAAGCGAA
33951 ACTTTTCTTG AAAGTCAAGA TAAATATAGA GAACAACAAG ATTCTGCTAA
34001 AAGTGTGCTG ATTTTAGAGA GTTGTGGTAA TTCTCTGTGA AGAGTTAGGT
34051 AAAATGGTGT ATCCTGGCTA TTTAAATGTT TTCTACTTAA TTAAAAATGT
34101 TACTGCTTTA ATTTATTTAA GATGCCAAGG GGAAACTTAG AAGTTGTTCA
34151 TCATCGAGCC CTGGTTTTAG CTCAGATTCG GAAGTGGTTG GACAAGTAAG
34201 TGCCATTGTA CTGTTTGCGA CTAGTTAGCT TGTGATTTAT GTGTGAAGAC
34251 AATAAGTATT TTATTACAAT TTCGAGAACT TAAAATTATG AAAAGCCCTC
34301 ATTACCTATA TCATCAATCA GATTCTTAGA GGCTCTTTTT TTTTTTTTTA
34351 ACTTTTTTAC TTTAATGCAG TATTTTGTAG TGGAGATTCC TAGCAGAAAG
34401 AATCGTGACA CTCATCATAT AAAGGAGGGC TTCTCTTAAC CTGAGGGAAC
34451 ACATGTGGGT TTTAGGTGGC CTGTGAACCC AGGGAGATTG TACACACCAA
34501 ACCTTGTCTT TGTGTATTTA TTCAAGTAGA AAGCCCACAG CTTTCAATAG
34551 ATTTACAGCG GGGCCTATGA CCCAGAAAAG CCTGAGCTAC TCTTGTGAAG
34601 GAAATGACTG ATTTTCTGAA CCTATTTGGA GGAAACTTTG TATTGGAAAG
34651 ATCTATACTA ATGTTTTGTT TAAAAAGTAG ACCTGAATTC CATGATGATT
34701 TTCTTTGTTT TTTTTTTGAG ACAGAGTCTT GCTCTGTCAC CCAGGCTGGA
34751 GTACAGTGGC GCAATCTCGG CTTACTGCAA CCTCTGCCTT CTGGGTTCAA
34801 GCAATCCTCC CACTTCAGCC TCCCGCATAG CTAGGATTAC AGGTGTGCAC

34851 CACGCCTGGC TAATTTTTTT TTTTGTATTT TCAGTAGAGA CAGGGTTTCA

34901 CCATGTTGGC CAGGCTGGTC TCAAACTCCT GACCTCAAGT GTTCTGCCCA

34951 CCTCGGCCTC CCAAAGTGCT AGGATTACAG GTGTGAACCA CCGTGCCCGG

35001 GCTTCTGTAA TGATTTTCTG TTGTATGTAT GTGAAGATGT AGTTCTCAGA

35051 CAGTCATGAT GACTAAATTA CACCTTTTAA GAAGGTAAAT GAATGTGGTA

35101 CCTGATTTTT TTATTCTGTA ATTTCAGAGT AGAAATCCAG TGATAGTAGC

35151 TTGGCATTGG GGCTGTAATC TGATTATAAC TGGTTTGTAT CATAATGAAA

35201 ATATGCTGGG CCCATGGAGC TCAGTTTTTG TGAATATCTT TTCTATTCTT

35251 TCTCTGTCTT CTCACAGACT TATGTTTAAA GAGGCATTTG AATGCATGAG

35301 AAAGCTGAGA ATCAATCTCA ATCTGATTTA TGATCATAAC CCTAAGGTAA

35351 CTTTCTAAGC TGTCATTTAC TCTAGCTTAC TTTGTACTTA AACTAATATG

35401 ATCTGAACGA AGATGTTTTG TCCTTTTTTT GGTAGGTGTT TCTTGGAAAT

35451 GTGGAAACCT TCATTAAACA GATAGATTCT GTGAATCATA TTAACTTGTT

35501 TTTTACAGAA TTGAAGTAAG TATTTTGAAT AATTCATGTG TATCTTTTCC

35551 ATAGTTTTCT CTCTTCTTGT TAAGGAAATC AAGCATAAAT AGCTAGAGAA

35601 GAAAAATTCC TTACTGTTCA TTTTTAAAAA TTGCTATAAC TCTTAGATGC

35651 CAGTTGGTTT TTTGCTCTTT TCCGTTCTTT TTAAAACAGC CTGTTTAAAA

35701 CTATGTCCTT AAAACATGTC ATTCAGAATT ATTATTTCAC TTGATTTTTA

35751 GGTATACATA TAAAACTACT TGTTTTTCCT AGGAGACTGA AATCAAATGG

35801 CATCTTTCTC TCTGATGATC TTTCCCCTCA ACTTTTTAAT GAAACACTTT

35851 CAAAATAGAG AAAAGTTGAG AGAATTGTCC AGTAAGCAAC CTATATATAC

35901 CCCACCTGGA TTCGCCAGTT TATATTTTTC TGTATACACA TTCTCATTCT

35951 CTATAATCTG TCCATCCATC ATTCATCTTG TTTGTAGACA AATTGCTAAG

36001 TGAGTTGTAG ACATCAGTCC ACTCTACCAC CTGTACTTCT CCTTGTATAT

36051 CATTAACTAG AGGGCATTCT TTGTGTATGG GTTGGTTTTG TTGTGTTTTT

36101 TCAGGTCATA TTTATCTACA GTGAAATGTC CAAATCTTAA GTGTGCCACT

36151 TAGTGAGTTT TGGCAAATGT ACACTTCATG TAACCTGAAC CTCTGTCAAG

36201 TTAGAGGGCA TTTACTCCTT TTCAGAAAGC TGCTTCAGAT TCCTTTCAAT

36251 CAGTCCCTGT CCCATTCCCC AGGCAACTAC TCTTCTGAAT TTTTACCAT
36301 AAATCAGTTT TGCCTGTTCA AGAACTTCAC CTAAATGGAA GCATACAGTA
36351 TTACTCTTCT GCATAAAGCT GTTTTCATTC AGCATATTGT CTTGAGATTC
36401 ATCTGTGTTT TTATATGTAT CACTAGTTCA TTCTTTTTTT ATTGGTCAGT
36451 AGTATGCCGT TGTGTAAATA CACCACTATT TGCTTATTCA TTCCCCTGTT
36501 GCTGGACATG TGGATTGTAC TACCCTGTTT GGGGCTAATG TGACTAAAAC
36551 ATCTACAAAC ATTTGTATAA GTCTTTTGTG GACATGTTTT ATTTCTCAAT
36601 ATTTTTATAA TTCAACTCTT TTCCAAAAGT CATTTTTATT TATCATCATC
36651 AGCATGCCAG GTGTATGTTA GTAATTTGAT CGCTGGGCTA CATGTTCTGT
36701 TGATGACCAT TCCATACACA CCTGTTCTTA GAGAAGAAGA TGTCACGAAG
36751 ACCATGTACC CTGCACCAGT TACCAGCAGT GTCTACCTGT CCAGGGATCC
36801 TGACGGGAAT AAAATAGACC TTGTCTGCGA TGCTATGAGA GCAGTCATGG
36851 AGAGCATAAA TCCTCATAAG TATGTATGCT GTCACCAGGT GGCATCCTTT
36901 GAAAAACCGA AGTGTGTAGT TGTCCTTGTC CAGCCTACTT ACCTTTCTCA
36951 TTCTGGTGTT CTTCACTTAT TACCTCAGAT ACTGCCTATC CATACTTACA
37001 TCTCATGTAA AGAAGACAAC CCCAGAACTG GAAATTGTAC TGCAAAAAGT
37051 ACACGAGCTT CAAGGTAGAG ATCCGCTCAC AGAGAAAGTG CTTAAGGTGG
37101 CCGTGACTGC TACTAGTCTT CTGCAGGTGA CAATCACCAT GTCATTGCCA
37151 CACCACAGAT TTAACATGTG ACTTTTTAGT TGCCATTTTA AGACCCTTGT
37201 CAGTTTTTTT CAGTGCTGCC CTCTAAAGCA TATATAAAAG TATCAGAAGT
37251 ATATATTCTT CTGATGTCCA GTTCTATTGA GAAAAATTTA TTGTCTTTTT
37301 GGTTATGTTG TTAGGTCTGT GGATTTTTTC CCCAAATGAT TGTGTTCTGT
37351 TTTGTTTTCT AAACACTGTT AGGAAATGCT CCCTCTGATC CTGATGCTGT
37401 GAGTGCTGAA GAGGCCTTGA AATATTTGCT GCATCTGGTA GATGTTAATG
37451 AATTATATGA TCATTCTCTT GGCACCTATG ACTTTGATTT GGTCCTCATG
37501 GTAGCTGAGA AGTCACAGAA GGTATGTGGA GTTCTTACTT TTATGCCATT
37551 TGGTTCTTGT TTATATAATG ATAGTGTGAA ACCCTGCTTC TGGTAGTGCA
37601 GTAGCTTTTC TGCTATCACT CTGTGAGTGC AGGGCTGGAG ACAGATCTGT

37651 GAGTTTCTAG GGCCCACATT CCTAAGCCCC TGTGCTTATG AAAGTGTTTT
37701 GATTGTGAGG TTGAAGAAGT GAAGTAAAAT TGCATGGCTT TTTTTTGTTT
37751 CTTTTTTTTT GAGACGGAGT CTCACTCAGT CGCCCAGGCT GGAGTGCAGT
37801 GGTGCGATCT CGGCTTACTG CAAGTTCCAC CTCCCGTGTT CACGCCATTC
37851 TCCTGCCTCA GCCTCTCTAG TAGCTGGGAC TACAGGTGCC CATCACCACG
37901 CCCGGCTAAT TTTTTGTATT TTTAGTAGAG ACAGGGTTTC ACTGTGTTAG
37951 CCAGGATGGT CTCCATCTCC TGACCTCGTG ATCCGCCTAC CTCAGTCTCC
38001 CAAAGTGCTG GAATTACAGG TGTGGGCCAC CATGTGCGGC CTAAAATTAC
38051 ATGGTTATTT TTAAGATGAT GGGCATATGT GTGAGCTAAT TTCTTCTCTT
38101 ATAAAGGAAA TGTAACAAGT GGTTCATGTT CCACTCCGGT TCTTTCTCAC
38151 ATGGCTCTTT TTTCTAGTGG AGGGTGGGCA CATGGAGCAC AGAAGGCTCA
38201 TGGCCTCCTT TCCTATGTTG GTACATTTGC TATGATCAAA AACTTTGAAC
38251 ACCACTGGTA TGCATATTTT TTATTTATTT TTTTGCAGCC TCAGTCTCTT
38301 CCCCATGACC TCTCCAAAAA TGAAAATCGG ATCCTTCATC TCTCTGCTTA
38351 AAATACTTCA TGAGCTCCCA TTGTTCCGAG GATATAATTC AGAAGCCATA
38401 ATACTGCTTA AAAACCCTTC CTTGACCTGG CCTCTGTGTA TCTTTCCATT
38451 CTCACTTCTT GGTATTGTCT TTTTTTCCTC TGCCCATGGA GGAAAGACAA
38501 TGCTTTTGTC CCCCTTCCCT TGCCCCTCAC CACCACATGC CTTGGTGGGC
38551 AGCATTACTT CTGCCATCCA TGGGCTTTGA CTGCTTCCAC CCTCACCATT
38601 CCCCTGGCTA ATTCTCACTA ATCTAGGTTA AAGGATGCCA AGGTGGCCTC
38651 TTCCCAGTAA GCCATTCATG CTTCCCTCCA GGGACTGGGT GAGGTGACCC
38701 TCCTATATGC TTCTGTTGCA CACAGTGCCT ACCCCTGCAG ACTACAGTGT
38751 GTCTTTATCT AGAGTGCGGT ATTTATTTAT TTATTTTTGA GACAAGGTCG
38801 GGCTCTATCA CCCGGGCTGG AGTGCAGTGG CACCATCTTG GCTCACTGCA
38851 ACCTACGCCT CCTAGGCTCA AGCAATCTCA CCTCAGCTTA CAGGCGTGCA
38901 CCACCATGCC TGGCTAAGTT TTGAATTTTT TTTGTTGAGA CGGGGTTTCG
38951 CCATGTTGCC CAGGCTGGTC TCAAACTTGT GAGCTGAAGC AATCCATCTG
39001 CCTCGGCCTC CCAGAGTGCT GGGAATGAGC ACTTAATTAT TTGTTGTCTT

39051 GGGTTTTCTT CCTATGTTGT TCTTACATGT ATTTATCCTG TCAGCCCAGG
39101 GAAATTGCAT TAAAAACAGG AAACACCTCT CCATTAGGAA GAAAAACAAT
39151 TTGCTTACAG GGCATGGCAT AGAGCTGGAG ATGATAGTGC CAATAAATAC
39201 TAGGTTGGCA GGGTCTCAGA GTTTTGTGTC CAACTCAGTA TAATTTTATG
39251 TTTGTTTTAA TGTGATCATT TCAGGAGAGC ATGGAATGTC ATGAAAACAG
39301 CACCAAGAGC AATGTCTTAG ACTTTTAGGA GAAACTTAGA TGCATTTGTT
39351 GAATATCTTC TAGACTGAAA CCTTATTTCC CTTATTAGCC TATGAAATAA
39401 ATGATACTGT GAGACTTAGT TAAGGAAGTT ACTATTATTC CAAGTGTAAC
39451 TTATTAATAT CCGTATGTGA AAGCATTTTT GCCAAAGCTT GTTTGATGTT
39501 CAGCTGACCC TTGCACAACG TGAGTTTCAA CTGTGCGAGT TTGAACTGTG
39551 TGGGTTTATC TAAATGTGGA TCTCTCTCAA ACACAGTTGG CCCTTTGTGT
39601 CCACGGCTTC TGCATCCACA ATCAGTGTGG ATCAAAAGTA CAATATTTGC
39651 AGGATTTGAA ACTTGCAGAT ACAGAGGGCC AACATTTTGT GTATCCAGGC
39701 TCCATGGGGT CAAATGTAGG ACTGGGGTAT GCTTGGATTT TGGTATCCTT
39751 GGGGTGTCCT GGAACCAATT CCCCATAGAT ACTGGGGGAC AACTGTAGTT
39801 TGATTTTATA TATTATATAA TATGCAGTTA ATATATAATA CACATTTAAA
39851 AATTATGTAG CTTTGGGTTT ATTGCTATAT GTAAATGCTA GTTTCTATTC
39901 CTATATATGA ATATCACAAG TAATAAAGTT CTCATTAATC ATTTTTTTAG
39951 GATCCCAAAG AATATCTTCC ATTTCTTAAT ACACTTAAGA AAATGGAAAC
40001 TAATTATCAG CGGTTTACTA TAGACAAATA CTTGAAACGA TATGAAAAAG
40051 CCATTGGCCA CCTCAGCAAA TGTGGTAAGT GTGGGGATTA GTATGTTTAT
40101 CTCTACTTCA GATCTTCTTT GGAACTAGGC AAGGTATAAA TTAAACTGTT
40151 AGTTTAGACA GTGACTGATT TCACTTCCCA CTCCTGAAAA CTCTAACAAT
40201 TATGTATGCT CACGTTATTT TGTCCTGTGT TCTGAAAAGC TGAAGGTAAT
40251 CACTTTTAAT GAACTGGAGG AGCTCCCTAG GTAAGAACGT CAAGTAGATC
40301 CTTTTTTGGT TAAGAATGAG CACCTGTGAA GTTAACTTCA GTGTCTCAGA
40351 ATCAAAATTG GTTGACAGTT CTTCCTTCTC ATGCTGTTTG CAGACATGTC
40401 AGGGAAACTC TGCTTGTCTG GAGAGAGTGA TGAGGCCACC TCCCCGTGCC

40451 CTGCAAGACG CAGTTTTAAT TGACAGTGAT GGGGTGCCAG TTGTTCTTCC
40501 CATGCTGGAA CAGTTGTGAT TCTTTACTGA GGACTGATGG GGGAAAGGAA
40551 GAATCACCTG GGGTGCATGT TAAGCCTTCA GCTGCTGGCA TCCTTGGAGA
40601 ATCTGATTCA GGTGGTCTGG GATAGGACTG AGGCGTGCAT GTGTCTAATA
40651 AGCTTCCCAG GTGATGTCTT TTCAAGGAGG CTGAGAAAAC ACTGGGCTGG
40701 AAAGCTGGGA CTCTTAAGTA GGATGCTGAT CCCAATCAGT GCTGCTCTTG
40751 CCTCAGAATC TGCAGTGGTG CTCATTAAAA ATTCAAATTC CAGGATCCCA
40801 TTCTTCAGAT TCTCTGATTA TTTAGGTCTT AAAAAGTTCC TCATTTATTT
40851 TGTTTGGTGA CCATTGGTAT AAATGAAGTC CATTATGCTT CCCATGTCTT
40901 AAGCCTGTCT TTGTGTGAAT CTTTTTCCTG CAGGACCTGA GTACTTCCCA
40951 GAATGCTTAA ACTTGATAAA AGATAAAAAC TTGTATAACG AAGCTCTGAA
41001 GTTATATTCA CCAAGCTCAC AACAGTACCA GGTATGTGGT ATGTGAAAAT
41051 GAGGCTCTCC TGGTTTTGCT TTTTGCTTTA GTAGGAAAGG AGTGAGGATC
41101 CTAAGTTCAT AACACCATCC TTGGCTTCAA AATTTATCTT AAAACTAATT
41151 AGCCTCAATT TGAACTTCTT ATCTGGGAGA ATGGTCCTGA CCTGTTCTCT
41201 GATTCCTCAT CTGGAATACC ACAGCACCTT CCTCGTGGGG TTCCCTGCTT
41251 CTTTCCCACC CCTCCTCTAG CCCAACCTTA CTGCTGTAAG TCTGATTATC
41301 CTAACAAGTA CAGATCTTTC CCATATATTT CAGCATAAAG GGAAATTTTT
41351 GTTTGCTTGA AAAAGCATCC CTTTAGCTTT TTTTATATAC CACACACTTT
41401 GCTTCTAAGT TAAATGTGTT ATATGATCCT CTTAACAGCC TCATAGGGTG
41451 CTGTACACAA TTTGTAGATG AGGAAGCAAC TTGCCTGAGG ATCCAGAGCT
41501 ACAAAGTGCT GGACCTGGGA TACAGAGCCC AGGCTGCCTG ACCACCCTGC
41551 CCATGCCATT AACCACCACT CTACCATGCC ACCAGCATCA CCATTTTCAG
41601 TTTGTCCTCA GACAATATAC ACATCTTTCT TTGATCAAGC CCCTGCCAGC
41651 TTCTTTAGCA CCAGCTTCTG CCACTGTCCA CATTCCCAGT TACTTGTAGG
41701 TAGTTCTACA GATGTCACAT CGTGTGATTC CTCTGTCATT TCTCTACCCA
41751 CCAGCCTTCC TTTAGCCCCA TTTGTCCATC AGAACCCTTG GGTTACTCCT
41801 GAATGCCATT CCTGGACCAG GCGCCAAACA CTGAGCCCCC AGAGCAGCCT

594221_1

41851 GCCCTCGCCT TGGTGATTGC ATTTGTCAAA CTGCTGATTA GCTGGTTTGT
41901 CACCTCCACC AGGCTGTGGG CTCCTTAAGG GCAGGGACTC CATGTTGTAT
41951 TCCTCTCTGA ATCTCTGGCT AACATCCAGC CTGGAGAATC GAGGATTTGG
42001 CCAGTGGATA CCTCTTTGCC CTTGTTTTCT GTTCTCTTCC ACACTCTCTC
42051 TGCTCTAGTC ACACTGGCCG TCCTGTTACT CCTCAGACCT GCTATACACA
42101 TTCCTGCTGC ATGGCCATGG TGCCTTCTGT GCCCTCTGCC TGGTGCCCCC
42151 TATCTCATCA CGTGGTTTAT TCTCCTGACA GCCATTAGAG CTCACACTCC
42201 CTGAGAGCTG CAAGGAGACT GTCCTCTGTC CCTTTACTCA CGTTTGCCAT
42251 TATGCTATAG ACTATATTTT GTCCCTAAGT CCATCCTCTG TTACTATAAG
42301 AGCAGCAACT TGGTGGTGGT TCTTATATGG TTTTTCATTT GTTGGTTTT
42351 ATTTTTTGCC TTGCTGTAGT ATCCATACTG CCCAGAATGG TGCATATGTA
42401 GTTAAGAGTA ATTATTTGTT GAGTGAATAA ATGGCACATC CTCAGTAAGG
42451 TTTTGAATGA AAAAATGACT GTACTAACTG ATCAACTGTA AGATTTTCCC
42501 AGGTAATTCT TTCAAGGGAG TTCCAAGTAT AGGAACTAAG GCAGCTACAC
42551 TGGAGCTTTA GAGAAATGAT TGTCATATTT CCTCCTCAGT CCTAAATCTC
42601 CTCTTGTCAC AGGATATCAG CATTGCTTAT GGGGAGCACC TGATGCAGGA
42651 GCACATGTAT GAGCCAGCGG GGCTCATGTT TGCCCGTTGC GGTGCCCACG
42701 AGAAAGCTCT CTCAGCCTTT CTGACATGTG GCAACTGGAA GCAAGCCCTC
42751 TGTGTGGCAG CCCAGCTTAA CTTTACCAAA GACCAGCTGG TGGGCCTCGG
42801 CAGAACTCTG GCAGGTAAGT ACAATCATTT ATATGTTTAC ATCTACAAAG
42851 GTTTTAAAAA ATTTATTTCT TTTGTTTGGT AATTTTGCAA ATAAATTTAG
42901 GGCAGAATAC TCTGAGACAG TCTTGTTCTC ACTGATAAAA ATTAATTTAG
42951 AATGCTTTAA AGGATAAGCT ACTACAGCAA GAGTCCCAGA ATGCAGTGGC
43001 CCAATATGGA AAGAAGTTTA TTTCTCTCTC CCATAGGGAT TTATAGGCCC
43051 TTCCGTTGTG TGGCTCTGCA ACCTTTTAGG CAGATGGTTG TAGCTGGGTT
43101 ATCTCCACAG CTGTGGGGAA GGAAGGAGAG TGGGGAGAAG TTAGAATCAT
43151 GGTAAAACAT TTACCTTTAA GTTGGAAATG ACCTGGATGG AAGTTAAACT
43201 ATCACCTTCT ATTCCATCTC GGCCACGCCA TGTAGCTGGA TGGGCTGTGC

43251 CCTGTAAGAA GGTAAAGATG AATTTTTGGA TGGGTCCATT CTGTTATAGA
43301 CAGTAGGTTG TTGGAATAGC CAGGAATGAG GTGGGGAAAA TAAAAGGCCA
43351 AATGTCGAAG CATTCTGAAA GCAAAGGCAG TTTAGCTGCG TCAGGGACAA
43401 GGGTTGCCCG AACCAGAGGC GAGGCTGGTA CCAGGGGCTC TAGTACCAGA
43451 GTGGAGGAAA GGGTAAGGAC ACCTATGAAA AGAGATGAGC AGAAGCTCTG
43501 GTCATCTCAG CAGTGCTTGA AGTAAAGCAA TGACTGGTAT ATTTTTTTCC
43551 CTAACTTGTA AATATTGTTG AGATCTCAAA GAAAAAAATA AAAAGCAGTC
43601 CTAAAAAAAT TCCAAACTCT ATCCTGTTAA ATTTTGTTAA ATTTATGTAC
43651 CAGTCCTTCT TTGTCATTTG CAGTATTCTT TTTTTCTTGG GATTATACCA
43701 GTGTATGGGA TTATCACTTT TCTTTTTCTG GTTATTAGCC TTTCCCAAAT
43751 CCCTCCGTTT CCATGCTGGC CTCTTTTTAC AAATGTCGAG AATTCCTTAT
43801 TTCAGGCCTT TTAGTTATTC GTTCGGTCTC CATTGTTCCT TTCTGCTTTA
43851 GAAATTTATG ATATTGGTTG TTTATACCTT CTATCTCTGT TCTTGGATCT
43901 CTTCTATTCT TTACAGCTCT TAGCTTGCTA TTTCCCATGT CTTATGAGGG
43951 AGTATTTCTA GTTTTTCTCA GATGTTTAGC AAAAGTAGGT GGGGAGGGCA
44001 GTGGTCAAAG ATGTTTGAGA AATGTTACAC ACTGGAGTCA CTCTGTGTGT
44051 ACATTTAACG TAGGCAGTTT ACACAAGAGA GCAAAAGAAA GGTAACTATT
44101 TAAATAGTGG AGGTGATTTT ACCTACTTTT TTTAGTGATA TATGCACTGG
44151 AGTGAGCATG CAATGAGAGA CCGGAATCTA CCAGCTCCTT CGAAAGCCTT
44201 GGGTTCTCTG TGCCTCTCAT TGTGGTTTAT CTCAATTGGG CTGAGAGTGA
44251 TTCTAGGATC TAAAGACACT GCATGACTCA AACATAAGTC AGCTACCTCC
44301 ATCTAGTGCT CAACCAAAGA AATAGTGGTC TCTTACTGTT AAGGGACGAA
44351 GTGGTTTAGT GAGAGATACC AGGTCATTTT CCCATATACA TGCTTTGGAA
44401 GCATCTTTCA AGGCTAATTT TGGCTGTATA TGATTTTCAA TTCCTGTGCT
44451 AAATTTAGAT TCTAGCTGCC ATTTAAGATA GGACTCTGTG GTGTATATAC
44501 CTATTCCCTC ACAGAAATTC AGAAAGTACA TAGTTTCATA CATAATAAAG
44551 ACATATTAAA GAAGCACTTG AGCTAAAGTA TCTGTTTAAC TTTGTAGTCA
44601 ACTGCTGCTT ATTGTCTCTA CAGGAAAGCT GGTTGAGCAG AGGAAGCACA

44651 TTGATGCGGC CATGGTTTTG GAAGAGTGTG CCCAGGTAAA CTCAATTCCT
44701 CCCTTCTAAA CCCCCCAGTC AGCAAGAAAG GTCTTCTCAA TTGTATCTTA
44751 GTGATCATGA AAGTTAAAGG AACTGTGCAT AATTGTTAAG TCCAGAGATA
44801 GTGTTTGCCC CAGAGGTCTT ATCTTGCTGG CTTGACTTGG AAATCTAAAT
44851 TTAGTACATC TCTAAGTTTG GTGAGGTAGA ATATGAAGGT GCTCTACTTT
44901 AACATACCAC TGGTTTGACC TTGGTAGAAA GTACTTAATT ACATCTCAAG
44951 GTAGCTGTGC TTTTTAAAAT TGAGTTTGCC AAAGTAGAAA CAATGAGAAA
45001 GGACCATTAT AAAACAGGAT CATTGAAGGC TACATACTCT TGGCTTTTAC
45051 TCTCATTCTC CCTATTGGAA ATGTCTCTTT TACCTCAGGG ACCTGGAGGT
45101 ACAGCAGATT ATAAGGATAA GTACCCATAT GAGCATTTGG TAGTATTATA
45151 GGATTTATTA TGAAAATAAT AAAACTGCAG TAACACTGGC CACAGACTAA
45201 CAGTACACAG GTGCACAGTT GACACCAGGG ATTATTGCCT TGTAGAGTTT
45251 TGACCTTTGA TGAGAGAGTG TTTTTTACAG TTGTTACTGA TAGCACATTT
45301 ATGTAACTTA ATTGTGCTTT AAAAATATTT AATTGTCTCT TGTGTAATAA
45351 CAGTAAGTGA AAGACGATAA CTAAAATTTT ATATAATTAG ATCCTGGAGA
45401 GAATATTTGT TGGGTGATTG AATTGAAAAT ACCAGTGAAT GAAACATACC
45451 TAAAAGGGTA GATAGGTTGG GTTGGAAAGA TATACCACAT CGAGGGTTAA
45501 TTAAATGGAT AAGATGTCAT TATCTTTTTT TCTTTGTAAA GGAAGATTAA
45551 TGCATAAAAT TATTTTGTGT AATTTACATA CAATAAAATT ATGTGTTGTA
45601 CAGTTGTATA ATTTACATAT AATAAAGCTA ATTCACCAAT TTTAGATGAA
45651 GAATTCAGTA CATTTGGACA TATGTTTGTA GCTGTGTAAC CACCATTGCA
45701 CTCATGATCT AGAACATTTC TAACACCCCC AAAAGTTCCC TACTTCCCCT
45751 TTTGCAGTCA GCCTTCTCCC TCCACTGCCA GCCTTTGGCA AACTGATCAG
45801 TCAGTAAAGT TTCACATTAT CTAGAATTTC ATATAAACAG AACCATATGG
45851 TATGTAGTCT TTTTAATCTG GCTCCTTTCA CTCACATAGT GCATTGGAGA
45901 TGCATCCATG TTGTAGTTTA TTCCTTTGTA TTGCTGAATA GTATCCCATT
45951 ATATGTATAT GTCAGAATTT GTTGATTTAC CAGTTGATGT ACATTTGGAT
46001 TGTTTTCAGT TTGGGGTTAT TATGAATAAC GCAGCCATGA ACATTCTAGT

46051 GCAGGTCTTT ATGGGGACAG GAGTAGGAAT GCCACATCCC GTGGTAAGTG
46101 GATGTTTAAC TTTTTAGGAA GCTGCAGAAC TAATCTGCAG TGGCCGTATC
46151 ATTTTGCATT CCCCTCAGTG ATATGTGAGA GTGCTTCAGT GACTCCTATA
46201 CTCACCAACA CTGGGTGTAT TACTGTGACA CTAGATGTAT TATCTATTGC
46251 TACGTAACAA CTTACCTTAA AAGCTGGCAG CTTAAAACAA CAGACCCTAT
46301 TATCCCACTT TTTCAATGGG CCAAGAATCT TGGCTGGGCT TAGCTGGGGC
46351 CTCTGGCTCA GGGTCCTTTA CAAGGCTGCA ATTAAGGTAT TGGCCAGGGC
46401 TAGAGTCATC TCAAGGCTTG ACTAGTTTTT AATTTCATTT TCTAATGTTT
46451 TATTACTAGT ATATAGAAAT ATAGCTGAAG TGTTTTGCAG GGAGGCTGTA
46501 TAATTGACCT TGTATCCTGC AACCTTGCTA AACTCATTTA TTAGTTCTAG
46551 AAGCTCTTGG GTGTATTCTC TAGGATTTTC TACATCAACA AACATGGTTT
46601 CTATAAATAT AGTTTTATGT CTTTCTTACA ATCAATACTT TTTTCTATCT
46651 GTATTGCATT TTCTAGGGCT TCCAGTGTGG TGTTGAATAG AAGTGTTAAG
46701 AGTGAACATC CTTGCCTTTT TCCTGATATT GGAGAAAATT CACTTGTCTT
46751 TTAGCATTAA GTGTCATGTT TGCTTTTTTA AAATTTTATT CTATATTTATT
46801 TTATTTTTGA GACAGAGTCT TGCTCTGTCA CCCAGGCTGG AGTGCAGTGG
46851 TGTGATCTCA GCTCACTACA ACCTTGACCT CCTAGGCTCA AGCGATCCTC
46901 CCACCTCAGC CTCCTGAGTA GCTGGGACTG CAGGAACATG CCACCATGCC
46951 TGGCTAATTT TTGTATTTTT TGTAGGGATG GGGTTTTGCC ATGTTGCCCA
47001 GGCTGGTCTT GAACTGTTGG ATTCAAGCAA TTCGCCTGTC TCAGCCTCCC
47051 AAAGTGCTGG GATTACAGGC ATGAGCCTCC GTGCCTGGCC TGATATTTGC
47101 TTTTTTTTTT TTTTTTAATG CTCTCTATTG CAGAGTTGGC AAACTACAAC
47151 CTGTGACAAA TCCAGCATGC CACCTGTTTT TGTAAATAAA GCTTTATTGG
47201 AGCATAGCCA TGCTCATTAG TTTACATCTT GTGTATGGCT GCTTTAACAC
47251 TACAGCAGCA GAGTTAGAGT TGTGACACAG ATAGTTTGGC CCATAAGGCC
47301 TATATTTACT GTCTAATCTT TTACAGGAAA AATTTGCCAA TTCCTGCCCT
47351 CTTGGTTTGA GGAAATTCCC TTCTGTTCCT TGTTCTGAGA GTTTGTATCA
47401 TGAATGGGTG TTAAATTTTG TCAAATGCAT TTTCAACTAT GAAGGGTTTT

47451 GTTTTTAGAC GAGTGATATG GGGGACTAGG TGATTGATTT TCTACTGTTA
47501 AACCAACCTT GCATCTCTGG GTTCAACCCC ACTTGGTATT ATAGATTTAT
47551 TACCCTTTTT CTCTTGTGGC AGATTAGATC TACTAAAATT TTCTTGAGGA
47601 TTTTTGTGTT TGTGTTCATG AGGGATATTG TAGTTTTTTC GTGTCTTTGC
47651 CATGTTTTGG GTATCAGGAT AATGCTGCTG TCATTGAGGG GTGACAAAAA
47701 TGAGGGGTGG TGTCCTTTAC ACTTCTGTTT TCTGGAGGAT TTCATGTAGA
47751 ATTGGTATGA GAGTCTAGCT TATGGTTAAA AACCTATGTG TGATGTTTCA
47801 GACCTGACCA TAAACAATTA CAGACTTTAC CTAGGAGGCC ACATGGGGAA
47851 AAGCTGCCCT CCCTACACCA GACTTGGCGT ACTGCCAATG CATTACAGTT
47901 TCTAAAGGGA GTTGCAGTCA AGGACTCAGG GCCCCCTGTT AGTCATGCTC
47951 TTGTAACAGT ATTTGCATTG AGAGTCCTGG CACTTTCATT CTTAGGTCTC
48001 TCTATCTGAG GACATGGGCC AAGGTCTTCT TCAGGCACCT CTGCCAAGGC
48051 CTGTTTATGC AAGAAGGAGT GGAAAAACCT TGACATTTTT TTCCACTGTG
48101 ACTCACTACC CAGTACTTTT CCACCCTTAG CCCCCTTCCT TTGCACCCAT
48151 ACCCCCAAGA TCCATCAAAC TGCTAAAGCC TTTTTTTCCA AGCTCCTTCA
48201 ACAGTGAACC AACCCTCATG TCTGTGTGGA TCCAGCTGAC TCTTGACTAG
48251 TGAGTTGTTC CTTGGGAAAA AATGGAACAG AGAGAGTTGG TGCTTTCCCT
48301 GGTTTTAGCC TCTTGCTTAT ACCAATGCAA TGCCTGAAGG CTTAATTCAT
48351 TTTTGACTTG TTGCTTTGAT CAGCTACTCC AACACCTGAC AGCTCAGCTC
48401 TTTCTCCCAG CTCTTGGGAG ATATTTTTTT CTTTAAATGT TTAGTAGAAT
48451 ATACCAGTAA GGCCATCTCG GCCAGGAGTT TTCTTTAATG AAAGTTTTTC
48501 ACTATTAGTT CAGTTACTTT AGTAGACATT AACCTATTCA AGTTTATCTG
48551 TGTCTTCTGG AATGAGCATT GGTAGTTTAT GTCTTTCAAG TAATTTGTTC
48601 ATTTCATCTA AATTGTCAGA TTTATTGGTA TGAAGTGTTT ATAGTATTCT
48651 CTTATTTTAC TGTCCGTAGG GTCTATGGTG ATGTCCTGTC TTTCATTGTA
48701 GATATTGATG TGTCTTCTTT TTTCTGATTA TTCTGGCCAG AGGTTTATCA
48751 ATTTTATTGA TCTTATTAAA GAATGAACTG TTTCATTGTT TTTCTCTATG
48801 ATTTTTCTGT ATTCTATATC ATTCTTTTTT TATTATTTTA TTATTTTATT

48851 TGCTCTTTAT TTTTCTAGTT TCTTAAGGTG ATGGCTTACT TTTATTTTTT
48901 TCTTATTTTT TTCTTTTGTT GTTGTTGTTT TTTTAAAGAA ACAGGGTCCC
48951 ACTCTTGCTC AGGCTGGAGT GCAGTGGCAC GATCATGGTT CACTGCAGTC
49001 TCAAACTCCT ACATTCAAGC TGTCCTCCCC CCTCAGCCTC CAGAGTAGTT
49051 GGGATTACAG GTGCATGCCA CCATGCCTGG CTAATTTTTA ATTTTTTTTG
49101 TAGAGATGGG GTGTTACTAG TTGCCCACGC TGGTCTGAAA CTCCTGGCCT
49151 CAAGTGATCC CTCCACCTCT GCCTCCCAAA GTGCTGGGAT TCCATGTGTA
49201 AGCCACTGTG CCTGGCCAAG GTGATGGCTT AAAGCTATTG ATTTGAGATG
49251 ATTCCTTACT TTATAGTTTA AGCATATAAT GCCATAATTT TCCTCAAGCA
49301 CCGTTTTAGT TACGTTATAC AAATTTTGAA ATGTTTTGTT TTCATTTCCT
49351 AATTTCCCTT GTGATTTCTT TATTGAACCT TGGCTTATTT AGAAGTATGT
49401 TTAACTTGCA GATATTGGAG ATTTGCCAGC CATCTTTTTG TTATTAATTT
49451 CTACTTTAAT TTTGTTGTGA TTAGAGAACA TACATTTTAT TAATTTAAAT
49501 TTATAATTTA TTTTAATTTA TAATATGGTC TGTTTTACAG AATGTTGTGT
49551 GTGTATTTGA AAATAATATG AAAGCTACTA TTATTGGATG GAGTGTTCTA
49601 TAAATGTCAG TTAGATTAGG TTGATCATGC TGTTCTAGCT TTTTATATCC
49651 TTATTGATTT CCTCACTACT TGCTCTATCA ATGACTGGGA AAGTGTTGAA
49701 GTCTCCCAGT ATTTGTCTAT TTCTCCTTTG ATTCTACCAG TGTTTGCTTA
49751 ATGTATTTTG AAGCTCTGTT ATAGGTGCAT ACATGTTTAT GAGTATGTTA
49801 TAGATGTATT CATTTTGATA TCCTTCTTTC TCTGTTACTA TTCCTAATTC
49851 TGAATTTGAC TTTAATGTTA TTAATATAAT TCTTCCAGCC TTCTCTTGGT
49901 TAGTCTTTTC ATTGCATATC TTTTTCTATC CTTTTACTTT TAATCTAGCT
49951 GAATGTAGTC TTTATTTTGA AAGTGCGTTC CTTGTTGATA GCATTATTGG
50001 TTCTTTTTTT TTTTTAAATC TAATTTGACA ATCTCTGTCT TTTAATTGGA
50051 GGGTTTAGAC ATTTGCATTG AATGTGATTA CCAATATAGT TAGATTTAAA
50101 CCTACAGTCT TGCTGTTTGC TTTTTGTTTG TTTCATTGAT CCTTTGTTTC
50151 TTGTTTTTTT CTTTTTTTGC TTTCCTTTGG ATTTAGTATT TTTCATAATT
50201 CCATTTTACC TCCACTGTTG GCTTATTAGC TATACTTCTT CATTTCAGTA

594221_1

50251 TTTTAGTGGT TGCTGTAGGA TTTATAATAA ATATCATTAA CTGACCATAT
50301 CTTCAGATAA TCGTATACTA CTTCATATAT AGTGTAAAAA CCTTACAAGA
50351 GTATTCACTC CATAATACTT TGTTATTGCT TTTGCTTTAA GTGATCAATG
50401 ATTGTTTAAG GAAATTTTTT AATGACCTTT CATGTTTATT CTTTTTTTTT
50451 TTTTTCCAAA AGATTCAGTA TTTTCCGAGT TTTCAAAAAC TGCTGGCCAC
50501 TCAAAGTGGA TCAACAAAAA TTTAAGAGCT AAAACTGTAA AACTCTTGAA
50551 GGCTGGGCAC AGAGGTTCAT GCCTGTGATT CCAGCACTTT GAGAAGCTGA
50601 GGTGGGACAA TCACTTGAGC CCAGGGGTTT GAGACCAGCC TGGGTAACAT
50651 AGAAAGACCT TGTTTCTACA AAAAATAAAA ACACAATTAG CCAGGCATGG
50701 CGGTGTGCAC CTGTAGTCCC AACTTCTTGG GAGGCCAAGG TGGCAGGATT
50751 TCCTGAGCCT GTAAGTTTGA GACTGCAGTG AGCTGAGTTC ACGCCACTGC
50801 ACTTCAGCCT GGACAACAGA ACAAGACCCT GTCTCAAAAC CAGAACGAAA
50851 CTATAAAACT CTTAGAAGAA AACAGGGCTA AATCTTCATG ACTTTGGATT
50901 TGGCAATGGA TGGTTAGAAT TAATACCAAA AACACAATCA ATAAATTGAT
50951 AAATTGGATT TAATAAAAAT TAAGAACTTT TGTGTATCAA GGACATTGTC
51001 AAGAATGTGA AAAGACAGCA TATAGAATGG AAGAAGATAT TTGCAAATCC
51051 TATATCTGAT AAAGGTTTAA TATCCAGAAT ATGTAAGGAA CTCCTGCAGC
51101 TCAACAACAG AAAGCCAGTT AAATCAATTT TGAAATGAGC AAACGCCTGT
51151 AAACCCAGCT GCTTGGCAGA TTGAGACAGG AGGATTGCTT GAGGCTAGGA
51201 GTTCAAGACC AACCTGGACA ACATAGTGAG ACCCTGTCTA AAAACATTTT
51251 TTTAATTAGC TGGGTGTGGT GGCATATTCC TGTAGTCCCA GCTACATGGG
51301 AGACCGAGGC AGGAGGATCA CTTGGGGCCA GGCAGTCAAG GCTGCCGTGA
51351 GCTGTGATTA TGCCACTGCA TCCCAGCCTG GGCGACAGAG TGAGACCCTG
51401 TCTGAGAAAA AAAAAAAAAA AAGAACAAAA AAAAATTTAG AAGATTGCTA
51451 TTCTAGTCTA CTATTTTTTC AAAGGGTGGT CTTGTTAACA ATTCTGGAGC
51501 CCACCTAAAC CTGCTAAATC AAACTTGGTA GTAAAGCTGG GGAGATGGGC
51551 ATGTCTAACA GACGTTTCTG GTGGTTTTGA TGTCCAGGCG TGCAGAGAGA
51601 TGATGCTTAC CTTGTGTTTT GTCATTATTT TCAGGATTTA CACCCCTTCC

51651 TTGTCTTTTG TATCAATATT TATGGAGTCA TGAACTCTAG GATAGGCATG
51701 ATGTTGAGAA CTAGGAGTTC TCCCCTGGCC AGGGAGATAG AGGCAGGTCT
51751 GTGGTTAGTT TTGTAGTTGG CTGTGATGAC ATCTGACATG CTCTCTTCAC
51801 TTGTTGTCTT CTTCCTGTTC CCTTGTCAGG ATTATGAAGA AGCTGTGCTC
51851 TTGCTGTTAG AAGGAGCTGC CTGGGAAGAA GCTTTGAGGC TGGTAAGAAT
51901 CTTGTAAATC CTCTGGATGT TGGGTGCTAA GCAGAGAGAG CAAGCAAGGG
51951 ATTCCAGGTC AGTTGGAATC TCTTGTCTTC TGAGGTTCAT GAAATAAGTA
52001 GAAATAGGTC AGGTTCCTGG CTTAAGGAAA AGCGGTGTTA CTAAAATCAT
52051 TTTTATCATT CTTGATAATA ATTTGAAATA TTACTGTCTT TTACTGAAAT
52101 GAATTGAATT TCCTTGGCTG CCTTGTAGGA GGCCTGTTTT TCAGGAAAAT
52151 ATTCTGATTA CCTCTGAAAG TAATCCATGT CTTTCTAAGT ATCTTAACTC
52201 TCCAGTGACT AGAAGTTTTC CTTCCTAAAA TATCGTGTTT TTCCTTCTAG
52251 GTATGCAAAT ATAACAGACT GGATATTATA GAAACCAACG TAAAGCCTTC
52301 CATTTTAGAA GGTGAGGGTT CCATTTTAGA TAGAATTCCT CATTTGGAAG
52351 AAGGTGAGGA GAGAGAGATG AGAGAGTCTC CTCCTATTTA CTGTGTTTTC
52401 TTAATAATAT GTCATGTAGA CTCAATCAAA ATTACCACCT GGATATAATA
52451 TTTAATTCTC ACTAGAATTT TTAAATATGC TGAACTATTA AATGGTAACA
52501 AAATATTTAA ATGTTAGAAA CCTGTGATCA AATATGATTA AGAATCTTTG
52551 TATTTGGAAA TAGTAAACTT GAATATGAAC TATATTAGAT AATAATATAA
52601 CACTGATAAA TTTCTGGCAT TTAATAATCA TGTTGTGGTT ATATAAGATA
52651 ATATCCTATT ATTCTCAAGA GATAAATGCT GAAATATTTA GGAATGAAGG
52701 ATCATATCTC TGCCTTACTC TTAAAAGGTT CCACAAAAGT ATTAATGAAT
52751 GTGTGTATGC ATGCAGAGAA ACAGGAAGCA AAAAAATGTC AAAATGTTAG
52801 TAATTGGTAA ATCAAAGTGA AGGGTATATG TGTGTTCATT GAACTCTTAC
52851 AACTTTTATG TAGGTTTCAA CGTTTCAAAG TATTTTTTAA AAGTTACCTT
52901 TTCAAATGAA GTTTGTGGTT CTTAGAGAAC ATATGAATAT TACCAGTTCT
52951 AGAATACTCA GATGGTCACT GTGACCTCTT AAAAGCAAAG TGGAGAAGGA
53001 CATCAGTTTG ACTTATAGAA ACCTTAGGGA GTGGTTGATT TTAAGTTCTG

53051 CATTTTTATG CACATCTACC CTGTAAGTAA CGTCTGGCCT TTCTGACATT
53101 TACATGTATG CACATTCTTA CCTTGTCTGC ACCCCCTTCC TCCATCCTAA
53151 TTAAAACGTT GCTGGGGTAC TTTTTATGTC ATTCACTTTA GGTACCTCTA
53201 ACTGGGTACT GAAAACATCA TTCCTCATCT ATAATAATCT AACCAGCTCT
53251 TACTTAGATT TTCACCACTA ATGAGAACCT TTCTTAGATA AATGCCGATA
53301 ATTCATCTAC ATAGGCCCAA AACCTATTAA TAAAATGCAT CCTTGGATAG
53351 TAGTATTTTG CTTTTTTAAA ATGTATTCTA CTAGTGTTAT TTTTCTCTTG
53401 TGTATTTTTC CATTGGACAA TATTTATTAG ATACATTTTT TCCACATCCA
53451 TGGGCATTTT GATGGATGTT TAGCCAGAAA CATTTAGGTA ATTTTCTTCT
53501 TATTTTTGTT AACTGAGCTC CCCTCCCCTA CCCCCCCTTT TTTTGTTTGT
53551 TTGTTTTGTT TGTTTGTTTG TTTTGCCAAT CCTCCCTTGC TTTAGGTATC
53601 AAGTCTTCGT TCAGGTGATT TTACAAGTTC AGTGGTAGCG CATATTCTGG
53651 GATAATGTTG ATGAACTCTA AGATCTGGAA TCTCAGTCTC TAATTTGTTA
53701 ATGCTTATTA AGGAAAAAGA GCTCGCTTGG AAAACCTAGT AACCTCTTTC
53751 TTTTTGCTGA ATTTTAACCC TCCTTCACTG CTCCCCGCCT TTAGTTTTTT
53801 CTCTTTGCTT AAACCTCATG CTCAAACTAT TTTCCATTCT GCATCTCCAG
53851 CCCAGAAAAA TTATATGGCA TTTCTGGACT CTCAGACAGC CACATTCAGT
53901 CGCCACAAGA AACGTTTATT GGTAGTTCGA GAGCTCAAGG AGCAAGCCCA
53951 GCAGGCAGGT CTGGGTGAGT ATCTGCGTGA AGGCCATCGA CGTGCGGGGG
54001 CAGTGGGGTT GGGTAACGCC ACACATTGTC TAGATTGCTT GGTGATCCGC
54051 CTGCAATCTG ATTACTGTGC CATGGGCAAG TGTGAGGCTT CTGTGGAGCC
54101 CCTTCAGGGC CCTCTGTGTC TGTGTTTGTG TGTTGGTGAA GGGCAGGACC
54151 AAGCATGAAT GGGGAGAGCT CTGCCAGACA TTCCCACCTA CCCCCATTCA
54201 CCCAGAGCAG CTGACCACTT CCGTGTCTAA CAAAATGAGT TTCCTCATTT
54251 CCAGAAAAAA GTTCAGGAAA CTACTGATTT ACATTAGTAA TTACTGTATT
54301 TAATATTATC TCATTCATTT TGAGATCAAC TTTGCAATCA TTTTCATCCA
54351 TCCTTTGATA TGCACCAGTT GACTCTAGTT AGTTCATTTA CCGCCCTGAA
54401 AGTAAACCCA CACATTAGCA GGCAGTGTTT TCATCGGCTT CTGGTTCTTC

594221_1

54451 TTTTCTAGAT GATGAGGTAC CCCACGGGCA AGAGTCAGAC CTCTTCTCTG

54501 AAACTAGCAG TGTCGTGAGT GGCAGTGAGA TGAGTGGCAA ATACTCCCAT

54551 AGTAACTCCA GGATATCAGC GTACGTATCA CATTGATTCA GCACATTGAC

54601 TATATCCTGG GCATATAGGG AAAGTGGAAG CAAATAGATT GGTTTTCTAC

54651 TGGGACGGTG TAGTGGGAGT GGGGAGAATA TTCTTCAGCG CTGTGTGGAA

54701 GTTGTTCAGA CACTTTCCCA GCATATCTGA GACATTAAAC TTGGCATTGG

54751 AAGGTTTTCT TCCTCAGCTT TGTGGCTTGT GTGTTTTCCC ATTCCCCACG

54801 AGGCAGTTCC TCCCCTGAAT GCTCAGTTTA TATTAACATC TGATTTTATT

54851 TTTTGAACAA ATGTTGTGAC TAAATTATAG GCACTGAAAA AATGAAAAGA

54901 TAAGCTTCTT CAATTCAAAA TCAGGATTGG AAGAGACCAT AAATGTAAAA

54951 TAAGTCATAA CACTTTTACC AAATATAGTA ATTTGTCAGA AATATTTATT

55001 CAGCACTCAT ATGGTAGGTG CAGTAGATGT TACCAAAAAC TTATAAGGAG

55051 ATATGAGTTA TAAGAGTTTA TAGTCTTGCT TGGGATGTGT AAAGCAATGC

55101 AAGATTATAT ATTCAAACTG AATTTTGCTT TAGGAATTTA AAATGGAGAT

55151 CTGTGAAGTT GTGTGGGGTC ATCAGCAACT GCAAGAAAGT AGCCAGGCAA

55201 GGTAGCACAT GCCTGTAGTC CTAGCTACTC AGGAGGCTTA AAAATATCTG

55251 TGTAATTTCT AACAGGAGAT CATCCAAGAA TCGCCGAAAA GCGGAGCGGA

55301 AGAAGCACAG CCTCAAAGAA GGCAGTCCGC TGGAGGACCT GGCCCTCCTG

55351 GAGGCACTGA GTGAAGTGGT GCAGAACACT GAAAACCTGA AAGGTATATT

55401 CTCAGTCCTG ATGATGATTC CTGACCACAA ACAATAGTGA ATAGGCAGTA

55451 CAGACAGGCA GAGTTCAGTA GGTGATTAAG CTACCATTTT CCCAATTTGA

55501 GGAAAGATGA GAACTTTTAG CAGGAAGGGT CATGTCTGCA CACATTCCTG

55551 AAGCAGCCCT TCTTAGCTGG TAACTGAGAA GCCTTCCTCC ATTTGGCATC

55601 CCCCTAACTG AACTGGGAGA GATGCTTAAG CCAGGATAAA GAATTGTGGG

55651 ACACTGCTTT CTGCGTAGGC CCCCCAGCGT GCTTGATTTT CTTTTTGTAG

55701 TACATGTGTT TAATTATTCC AGCATTTGGG AAGAAAAAAG ATAATGTGGG

55751 AGAAAGGACC TGCAGTGGGA TCATAGAAAT TTTTGGCTTT GGATAGAAGC

55801 TATGTATGAT TCTGTCAATG GAGCTGGGAA TATAACTTAC CACTCTTTCA

55851 AATTTCTTCT CTCTAGATGA AGTATACCAT ATTTTAAAGG TACTCTTTCT
55901 CTTTGAGTTT GATGAACAAG GAAGGGAATT ACAGAAGGCC TTTGAAGATA
55951 CGCTGCAGTT GATGGAAAGG TCACTTCCAG AAATTTGGAC TCTTACTTAC
56001 CAGCAGAATT CAGCTACCCC GGTAAGTTTT CTCAGAGACG GTGTGCATTT
56051 TTTTCATCAT TTTCATGGGT TATTGTATTC ACACAATCTC CAAGTCAAAA
56101 AGTTTTCCTG TTCTTAAAAC ATAAGATGCC ATAGTTAAAT TATCTTAGCA
56151 TTTATGTGTA AGCTGTCAGT AAGATTTGAT ATTTGCCTGT AGAGTGACTA
56201 GTATACCTTG GCATAGGTTA AATGGACTGT CATTTTCCTT TCTGGATGAA
56251 GTAGCTGTCA TGGAGAAAAT GGGAAAGTCA CATGATTGCT CCTGGCCTTC
56301 AATGAGGTTG GAGTGGGGAG AGATGGGGGA AGATGGGGTC AGAGACGGCC
56351 TCTCACTTTC CTTTCAGAAC TCAGGGATGG GATCAGGCTT TAAAGGGACC
56401 CCAGGCAATT GCTTTTCCTT TTGTTTTATG AAAAATTTGA CTTGTCACTT
56451 CTATGTTGTT ATGATGGACT TTGCGGGTTG TGTTTAAGGC TGAATCAGCT
56501 TTGTATCGCA GAATTCTAGT ATATTGTCAT CTGTTTATTA TTTATACCTC
56551 TGTTCACTCT CTTATACTTC AAGTCTATTG TTAAGAGTTT TTATTTGGAT
56601 TCAAAAAGGC TGGTGTATCA GTCAAGATCT AGAAAGGAAA ACAAAAGCCT
56651 ATCTATTATT TTATCACAGA ATTTAATATA TGGATTTGTT AAATAAGTAT
56701 TAGAGGACTA AACAAGGCAA AAGGGAAATA CAGAGGAAGG ACATTGAGAT
56751 AGTAACTGTA GGAAGCAGCT TTACCCTCTA GCTGAGGGAA CAGGAGGAGT
56801 TGTTGGGAAT TATTAGAATT TAGAAGCCTG GAAGTGGGGC CCTGTAGAGC
56851 TGGCTCTTGA ACCTCTGAGA GGAGGGTGCC AGCCAGCTAA TCCTGGCATT
56901 TCTGAGGGAG CTGGTTCCAA GCGTACAGAA GTAAATGGAA ACTGGAAGGA
56951 ACAGCTGCTG CTGGGGGAAA AGCCAGCCGG TCGGGCCAGG TGTGGTGGTG
57001 GCTCACGCCT GTAATCCCAG CACTTTGGGA GGCCAAGGCA GGCGGATCAC
57051 CTGAAGTCAG GAGTTCGTGA CTAATGTGGC CAACATGGAG AAGCCCCGTC
57101 TCTACTAAAA ATACAAAATT ACCCGGGCAT GGTGGCGCAT GCCTGTAATC
57151 CCAGCTACTC AGGAGGCTGA GGCAAGAGAA TCGCTTGAAC CTGGGAGACA
57201 GAGGTTGTGA TGAGCCAAGA TCGTGCCATT GTACTCCAAC CTGGGCAGCA

57251 AGAGCGAATC TCCGTTTAAA AAAAAAAAAA AAAAAGCCAG CCAATCACGG
57301 AAGAAATCTA GAAATCTTTT GTTCATCCTC CAGCTTTGTA CTCCCCCTCT
57351 GGTGTTCACT GTAGGCAGGA CATGATGGGA AGCCAGCAGC AAGGAAGAAT
57401 ATCTTTCAGG TGCCCAGCCC CAGCACCACA AGCAGTGGAT AGAAGGGTGG
57451 GTTGGAGCTG AGAGATTACA AATCAGCTCA GTGTTTAGAA ACACATACGC
57501 TTATCATGTC TTGATTTCCT CATTTAGAAA TGGGCATAAG ACTTCTCTGT
57551 GTGCTTCAAT AGAATGCTTT GAAGGTTAAA TAAGAGGGTG TGTGTAAAAG
57601 CACTTTACAA ACCGTTGAAA TAAAAGCAAC TAGGAATCAG GGCCCCAGAA
57651 CTTCTTGAAT TTATTATAAT AGGTATTTCT TAGAAGAAAT GTGATCATCA
57701 TCTTCAAAAC TGTAGTACTT TTGAAGATAA TTGTTTTTGT TTTTTGAGAC
57751 AGGGTCTCAC TCTGTTGCTC AGGCTGGAGT GCAGTGATCA CCGCTCACTG
57801 CAGCATCCAC CGCCCCGGGC TCAGGTGATC CTCCCACCTC AGCCTCTTGA
57851 GTAGCTGGGA CTACAGGCGC ATGCCACAAC ACCTGGTTAA TTTTTCAAATT
57901 TTCTGTAGAG ACAGGGTGTC ACCAAGTTGT CCCCGCTGGT CTTGAACAAC
57951 TCCTGGGCTC AAGTGGTCTG CCCACCTCAC CTCTCCAAAG TGCTGGGACT
58001 ATAGGCATCA GCCACCATGC CCGGCTTGAA GATAATAATT TATAATACCA
58051 CTCCCATGAG TGATCTTCTC TTCTGATCAC ATATTCACAT TAAGGTCTAT
58101 TTTATTTTAT TTTTTTCTTG CTCTGTCACC CAGGCTAGAG TGCAGTGACA
58151 GTATGATCAA TCATGGCTTG GTGCAGCCTC GAATGCCTGG GCTAAAGCAG
58201 TCCTCCCACC GCAGTCTCCT GAGTAATTGG GACCACAGGT GCACACCACC
58251 ATGCCCAGCT AATTTTAAAA TTTTTTCCTA GACATGGGGA GAGGGAGTCT
58301 TGCTGTGTTG CCCAAGCTGG TCTTGAACTC CTGGCCTCAA GTGATCCTCC
58351 TGCCTTGGCC TCCCAAAGTG CTGAGATTAC AGGTGTAAGC CACCATGCCT
58401 CCCACATTAA GTTCTAAGAC ATCAATTTTA TGATTGTGGT TTTGATTGGT
58451 GAAGTATGGT TGTGGTATGT GCAGGATACC GTGAGTGACT TCTCATGGCA
58501 TTGCTCTTGA GAGTGTGCCA CCAAGGGTCT GCACTAACCA GGGGTGTGCC
58551 CAGAGGCTCG CTGCAGGCTT GAAATTCCTG CGGAGTCTTG TGTTTTACCT
58601 GGAGCACATG TGCACAGTTT CCATTCTGCT CCATAGTATG CACATGTTTG

58651 TATTTATTTC AACCTAAAAA TGTTTGTTTC CCATAACTCT TTGCGTATAA
58701 TTGATACTCT ACGTATTTGT AGCCTCTTTT ACTCTTTTCC CTTTCCTCAG
58751 GGAGTGGTTT GCTCATTTAG AAAAGGCCAA GATATATCAC TGTAGAGTTT
58801 CGTTTCTTTT CTTTTCCTCC ACCCCCCATC TTTACCTTGT TCTGGGAGAA
58851 AGGAGAATTA GAAGTCTGAG TTGCAGCTGG AGAAACTGGC AAATTAAAAT
58901 CACATTGGGA AAGAGAATTA CTGTGTTTCA CACCATACCA GTAGAAATGA
58951 CAGGCTGTTT TCTGCTGGTA GGGATTTGGC CTTTGGTATT GGCAGTCTTG
59001 AGAAGTATTA GATAATCTTT GCTGATACAG TCTATTTTCT CCTCAGGTTC
59051 TAGGTCCCAA TTCTACTGCA AATAGTATCA TGGCATCTTA TCAGCAACAG
59101 AAGACTTCGG TTCCTGTTCT TGGTTAGTAT TTTTTCTCAT TTAATATTAC
59151 AATACTAAGC AGAAGGACTA TCTTTCTGTA AGTATTGAGA AGATCAGCAG
59201 TATAAGGAGA GATTGGATAC AATTTTTCAC TACAAAAAAT TGACTACAAT
59251 TCTTCCTCAA TTCTAAGACC GCATCTTTAG TATGATCAGT TTCATGCTTC
59301 TAGCGGTGGG GGACCTGGTG CAGGAAAATC CAGCATGACC ATTGTATGTG
59351 TAATTTTTAA AAATATTTAT GTGGCATATG CTTGTTCATA AAGGCACACC
59401 ACAGTTCCAG TTTCAGTCTA AACTGTCTAC ATTTACATAT ACATCAAAAG
59451 ATTCTTCTGA AGCATCATTA CTGGCTATTG GCAGTTATGC TTTGCATCTT
59501 GGGGGCATTT TCATAAACCT TGCTTATGAG TGGGACCTTT TTATTATGTT
59551 TAGGATTGAC AATATAATTT GAAGGCAAAT CCAAAGAATA TTAGCATTTT
59601 ATACATATTT CCTGTTTAGT TATGCATGAA GTGTTTTATT TGTTGAGGGG
59651 AGATGATTCT CAATTAGATT ACTTATTTCC CTAAAAATTA AAAACCCTAA
59701 GCGCTTTCTT TTGAAAGTTG GTTAGAAACA TTTGATGAGT CAGCTTGGGA
59751 CTTTCAGTAT TTGCCCTTAC TTATAGTTGG ATCAATGAAG CATCTTAGCT
59801 TTGAAAAGTG AATGATAGTT TCTAAAATAA TTGGCAGTTT TAACTGCTAT
59851 TATTTGCATT TCTAGCATGT GACAAGCAAC TTTCTGAAAT TTTTTTTCAC
59901 CGAAGTGCTA CACTGTAATA GCATTTTGAT GACATTTGAA GTAGCCTGTG
59951 GGGATTCAAA TTAAGTTTGA CTTTAACAGC TTATGTTGCT ACCAGGAAGA
60001 ACAGCTACCT TCCATCCCAG CTAAACTCAT ACATCCAGAC TGTAACTACT

60051 GTATTCCTAG CTCCTCTTCT GTCTAGAGAA TGGCAAGGTT CTTTTGGTAT
60101 GCAGTTTCGA CATATCCACT TATTCCTTTT TTTTTCTTAA GTTTTTTCAT
60151 TTAGAAAAAA AAACAGATGG GGTCTTAATA TGTTGCCCAG GCTGGTCTCA
60201 GCCTCCTGGT CTCAAGTGAT CCTCCTGCCT CGGCCTCCCA AAGTGCTGGG
60251 ATTACAGGCG TCTGCCCCTG TGCCCAGCCC ACTTATTTCC CAGATGCTAG
60301 GAACTTACAT TAGACCTGAG GCCATTTGGT CATTGTTTAT TTTGTGCTGT
60351 AGTCCAATCC AGTTGTGATT TCTGCCTCCT GTGTTCCTCG TTGCTGGCCT
60401 GATGCTGACC TTCAGGTTAG GTCAGTCCCA TCATTCCCCA GGGTATTCTA
60451 GATGGCTTTC CCACTTCAAA GAGCACTTTC TTGTTTTCCA GCTGAGCCTT
60501 AAAGACACTC TGTAATATTT GAGAGCCCCT CATTATCTGA GTGTTTATTA
60551 TCATTACCCT TGTGGTTTCA AGGATGTATA GGAAAAGGTA AGTTCCTATA
60601 ATTCAAAAAT TGCCACTGAT GAACTAATCA CAAAATTAGT GCCACTCAAA
60651 TATTACTCAG CTGCCCCTCC CCAGCTAACA ATAGTTAAGT ATATTGGCAC
60701 ATCCCCACAA GTGAAATCAA TGACTTGATG GGTCATTTCT GATTGTTTCC
60751 TGCTTTGATG CAATACAATA TCATGCAGAT CAATTGCAAG TCTTGCAAAA
60801 ATTTAGTATT ACATAAAATA GATTAAAATG ATATTGGAAA AGTACTTGAA
60851 TCACAGCTGG GTTGGACTTG TTGCAATTGA TGACAAAATA AGTGCTTCAA
60901 ATGATTTTGA CTATCAAAGG ATTGAGAGAG GTCCTTAGAA AAATTGAAAA
60951 GCCCTCAAGT TATTTTTATA AAAATGGCCT TTTTTGTGTG CTGTGAAATC
61001 CACATATGGA AATGTGAAAT ATGTCATGTC CTGCTGTCAT ATAATTTGTC
61051 AGAATAATTA CTTTCTTGCC CAAAAGTCTG TACTTTGTGT TTATTTCAAG
61101 TTAAGTCTAG AATCAAATAT AGTTGTAGTT ATGCCTAATT TTAAAAAATG
61151 AGATAGAGCA CATTATTTTT GTAACTAGTT TTTTTTTTTT TTTTTCAGAC
61201 AGAGTCTTGC TCTGTGGCCC AGGCGGGAGT GCAGTGGCGC AATCTCGGCT
61251 CACTGCAAGC TCCGCCTCCC GGGTTCACGC CATTCTCCTG CCTCACCCTC
61301 CTGAGTAGCT GGGACTACAG GCGCCCGCCA TCACGCCCGG CTAATTTTTT
61351 TGTATTTTTA GTAGAGACGG GGTTTCACCG TGTTAGCCAG GATGGTCTCG
61401 ATCTCCTGAC CTCGTGATCC ACCCGCCTCG GCCTCCCAAA GTGCTGGGAT

61451 TACAAGCGTG AGCCACCGCG CCCGGCCTGT AAATAGTTTT TTTAAGATAA
61501 AGTCTTATTC CAACTTTAAT TGGAATTTAT GAAATACCTT GTTGATAGTG
61551 AATTTATTTA AGTAGCCTTT TTTCAGTATT GATATTCTTA TATCTTTATG
61601 GCACCATTTA GTGGAGAGAA ATGTAAACAA ACATAAAGAT GTAGTATTAA
61651 ATCATAACTG CATAAAATTA ACTGTAGTAT GTACTGCACT ACTGTAATAA
61701 TTTTGTAGCT ACCTCCTGTT GCTATTGTGG TGAGTGAGCT CAAGTGTTAC
61751 CAATATCTGC TTAAAATGCC ATGTGCCGCT AACCATCTCC ACATGAGCAG
61801 CACATGAGAG TCTCCATTAA TTGCATATGG CAGCGAAAAG TGATCTCTTG
61851 CATTGTCGTG TATTTTTTAT CACGTTTAAT GTAATATCGT AAACCTTAAA
61901 TAACACCATG AGACCTATAG GAAGTACCAC AAGTGTTGCT CCCAGGAAGC
61951 AGAGAAAAGT CATAACATTA CAAGAAAAAG TTGACTTGCT CGATATGTAC
62001 TATAGATTGA GGTCTGCAGC TGTAGTTGCC CACCACTTCA AGATAAATGA
62051 ACCCAGTGCA AGGACTATTA TAAAAGAAAA GGAAATTTAT GAAGCTGTCA
62101 CTGCAGTTAT GCCAGCAGGC ATGAAAACCT TGTACTTTTT GCAAAATACC
62151 TTTTTATGTT GTATTGAAGA TGCAGCTTTT ATGTGGGTGC AGGATTGCTA
62201 TGAGAAAGGC ATACCTATAC AACTATTATG ATTTGAGAAA AAGCACAGTC
62251 ATTGTATGAG AACTTAAAGC AAAAAGATGA AGGATCAAAG CTGGAGAATT
62301 TAATGCCAGC AAGGATGGT TTGATAATTT TAGAAAGAGG TTTGGCTTTG
62351 TAAATGTCTG GATAATAGGA AAAGCAGCTC CTGCCATCCA GGAGGCAGCA
62401 GCAAAGGCAG TCAGGTTTAT GATCAGGACT GCCCTTATCT GTAAAGCTGC
62451 TAACCCCCGA GCCTGGAAGG GAAAAGATTA ACACCAGCTG CCAGGCTTTT
62501 GGTTGTACCA TACAACAAGA AGGCTTGGAC AAGGAGAACA CTTTTTCTGG
62551 ATTGGTTCCA TTGTCGATTT GTCCCTGAAG TTAAGTAGTA TCTTGCCAGT
62601 AAGGGGACTG CCTTTTAAAG TTCTTTTGAT ACTGGAGAAT GCCCGAGGCC
62651 ACCCCAAACT CCATGAGTTC AACACCGAAG ACATTGAAGT GATCTACTTG
62701 CCCCCAAACA CACATCTCTA ATTCAGCCTC TAGATCAGGG TGTCATAAGG
62751 ACCTTTAAGG CTCGTTACAA ACAGTACTCT ATAGAAAGGA TTGTCAAATG
62801 TATGGAAAAG AACCTTGACA GAACATGAAA GTCTGAAAGA ATTACACCAT

62851 CAATGATGCC ATCATTGTTA TAGAAAAAGC TGTGAAAGCC ATCAAGCCCA
62901 GGACAATAAA TTCCTGCTAG AGAAAACTGT GTCCAGATGT GCATGACTTC
62951 ACAGGCTTTA CGACAGCCAA TCAAGGAAAT CATGAAAAAG ATTGTGGATC
63001 TGGCACAAAA AAAAAAAAAA AAAAAAAAAA TGGTGCATGA AGGATTTCAA
63051 GATAGGAATC TTGGAGAAAT TCAAGAGGTG ATAGACATCA CACCGGAGGA
63101 ATTAACAGAA GATGACTTGA TGGAGATGAG TACTTCCAAA CCAGCGCCAG
63151 ACAATGAGGA AGATTACATA AAAGAAGCAG TGCCAGAAAA TAAATTGACA
63201 TTTGTTCCAA AGGTTCCAAT TATTCAAGAC TGCCTTTGGC TTCTTTTACA
63251 ACATGGATGA TTCTATGTTA TGGGCACTGA AACTAAAAGA AACTGTGGAA
63301 GGATTGGTAC CTTAGAGAAA TGAAAAAGCA AAAACATCAG AAATTATGGT
63351 GTATTTCTGT AAAGTTAGTG ACACTGAGTG TGCCCACCTC TCTTGCCTCC
63401 TCTTTAACCT CCCCTACCTG TTTCATCTCT ACCACCCCTG AGACAGCAAG
63451 ACCAACCCCT CCACTTCCTC CTCTACTTCA GCCTACTCAA CGTGGAGATG
63501 ACAAAGATGA AGACCTTTAT GATGATCCAC TTCCATTTAA TGAATAGTAA
63551 ATATTGTTTT CTTTATGATT TTCTTAATAT TTTCTTTTCT CTAGCTTACT
63601 TTATTGTAGG AATGTAGTAT ATAATACATA TAACATACAA AACATTTGTT
63651 AACTGACTTT TTATGCTGCC AATACACTGC CGAACAACAG TAAGCTATTG
63701 GTACTTGAGT TTTGGAGATT CAGAAGTTAA ACATGGGGCC AGGTGTGGTG
63751 GCTCACACCT GTAATCCCAG CACTTTGGGA GGCTGAGGTG GGTGGAACGA
63801 GACCAGGAGT TTTGAGAGTA GCCTGGGCAG CATGGTGAAA CCTTGTCTCT
63851 ACAGAAATTA GCCAGGTATG GTGGTGTACA CTTGTAGTCC CAGCTACTTG
63901 GGAGGCTGAG GCAGGAGAAT CGCTTGAACC CAGGGGGTCG AGGCTGCAGT
63951 GAGTCATGAT CGTGCCACTG CACTCCAACC TGGGCAACAA AATGAGACCC
64001 TGTCTCAAAA AAAGAAAAAA AAAAGGTATA TGCAGATTTT TGACTGTGCA
64051 GGGGGGTCCG CACCCATAAC CCTACATTCA AGGATCAACT GTAATTTTTC
64101 ATGCCTGCAT GGCTCATATG TACAGATTTA CTGCTGGAAG TTTATCATAA
64151 ATAATGCTGA AAAAGAAAAT CCTTATATAT ACATATTTTC TCCTATCTCT
64201 GCTTGCAGTA TATGATTCCT GGTTAGAAAA GAAACTTAAC AAATCTAAGT

64251 GAAAGAGTGC CTGGGAGTTT TAGGTTACAA TGACAGAATC TTTTCCTAAC
64301 CCTCTCTCTC CATTCACTTT TTTTAAAGCA GGGGCATCTT TATTGATCAA
4351 CATGTTTGTC GAAgtTTCAT CATAAAGTAG TTCCTGTCCA TTAACTTCAC
101 TTACTGAATA TGTGCTATCA CATTTTGCTA TTCCTTAAAA ATTGAGCTAG
51 ACTTTACATA TAGTGAAATG CAGAGATTTC AGGTGTACAA TTTGATGAGT
01 TTTAATAAAT GTATACAGCC ATGTGACTGC TGCCACCACC CCTCCCACCA
55 GTTTGAAATA CAGAACATTC TTCCACTTTG AATCACTGGG TGAGCATGCC
01 TGAGGTTGAA ATGCAGTCCC TCCTCTCAGG GCGGGGCCTC CAGGTTGTGT
51 TTGCTCTGAC CTGGAGGTTG CAGGGGTAGC AGACACATGA ACTCTGGCTC
TGATGGTCTT ATTGCTGCAA ACTCCACCTG CCTAGTTTGT TTAGTTTAGA
CTGCCT CAGCGCCCTC CAACAAGAGT ATGTCTGTCA CAATTTCCCT
TCCTTTCTTG CTTTTAGATG CTGAGCTTTT TATACCACCA AAGATCAACA
51 GAAGAACCCA GTGGAAGCTG AGCCTGCTAG ACTGAGTGAC TGCAGTTAGG
4901 AGGGATCCGA CAGAGAAGAC CATTTCCACT CATTCCTGTT GTCCTACCAC
54951 CCCTTGCTCT TTGAGGGCTG GCTATTGAGA ACTGGAAAGA GTAAAATGAT
65001 AACTTACCTT AGCATTGCCA AGAACTTCAG CAGACAACAA GCAATTCTAT
65051 TTATTTTATG TTGTGTATAC ATCTTGATCA TTAGCAAGAC ATTAACGCTTT
65101 AACCATTATG GCACCATTTT GTGAGAATGA TTGTTCTTTC ACTTGGGCTG
65151 TTTGAGAGCA TAATTATGGT AATCATGAGA TTAATGTTTC ATGATTTCTA
65201 CCTCCAAAGT GTGAAGACAA GTAAAACAAT GTTTCTAAAT TGTCTTATTT
65251 TGTTGGCGGA GAAGATTACA ATGGCTATTA GTGCTACATT TGGTCAAATG
65301 TAATCACTTA AATAGCTTCT TGTCACCTTA AACTAAAGCA GAATAAAAAG
65351 TATCCTTTGA AATTATAAGC CCTCCTTTGC TGACAGCTAT TATTTTGTAA
65401 CATCTTACCA GGTCATGTGC TTTCAGTTAT AACTGGGCTG AGCCTCCTAT
65451 AATTACAATG TCTATAGGGA CTGTTTTACT GCCTGTGTAT TTTCTGCTAG
65501 AGAGTTAGCA ATGTTAGAGC TAGAACAGAT TAGAATTTCT AAACAGTATC
65551 ATGCACAGTT GGTGTGAGTG ATCAGTGTGC ATTGTATGGC ATGCATGGTT
65601 GTGAATTATT CTCTGTTCTC CAAATACTGT TTCTTTAACT CAGATATTTT

65651 TGTTAGTGTC TAGGCCACTT CATTTATTTT TCGTCATGGT ACTTTACTGA
65701 CTTCTCTTTA TTCAATTCTC CACGCCCTCA CCAAAAAAAA CTGTCTCAAA
65751 ATGAGAATAT TTTATTTTCA TGGTGAGTCT AGAAAACGCC CACTTCATTC
65801 TGATTAAAAA TTCTTCCATG TTTTAAATAT CAGAACCAGA CCTTTCTTAC
65851 TGTGTATCTT AGCCCATTTG TGTCTCTATA ACAACAACCA GCTTTCAAAG
65901 GAACTAATAG AGTGAAAACT CACTCATTAC CACGAGGATG GCACAAGCGA
65951 TTCACGTAGG ATCTGCCCCT GTGACCAAAA CACCTCCCAT TGGGCCCCAC
66001 TTCCAACACT GGTGATCACA TTTCAACATG AGGTTTAGGG AAACAAATGC
66051 CTAAACTACA GCACTGTACA TAAACTAACA GGAAATGCTG CTTTTGATCC
66101 TCAAAGAAGT GATATAGCCA AAATTGTAAT TTAAGAAGCC TTTCCCAGTA
66151 TAGCAAGATG TTAACTATAG AATCAATCTA GGAGTATTCA CTGTAAAATT
66201 CAACTTTTCT GTATGTTTGA ACATTTTCAC AATCTCATAG GAGTTTTTAA
66251 AAAGAAGAGA AAGAAGATAT ACTTTGCTTT GGAGAAATCT ACTTTTTGAC
66301 TTACATGGGT TTGCTGTAAT TAAGTGCCCA ATATTGAAAG GCTGCAAGTA
66351 CTTTGTAATC ACTCTTTGGC ATGGGTAAAT AAGCATGGTA ACTTATATTG
66401 AAATATAGTG CTCTTGCTTT GGATAACTGT AAAGGGACCC ATGCTGATAG
66451 ACTGGAAATA GAAGTAAATG TGTTTATTG

FIGURE 7

```
   1 ccagtgctgg ggctgcctag ttgacgcacc cattgagtcg ctggcttctt tgcagcgctt
  61 cagcgttttc ccctggaggg cgcctccatc cttggaggcc tagtgccgtc ggagagagag
 121 cgggagccgc ggacagagac gcgtgcgcaa ttcggagccg actctgggtg cggactgtgg
 181 gagctgactc tgggtagccg gctgcgcgtg gctggggagg cgaggccgga cgcacctctg
 241 tttgggggtc ctcagagatt aatgattcat caagggatag ttgtactgtt ctcgtgggaa
 301 tcacttcatc atgcgaaatc tgaaattatt tcggaccctg gagttcaggg atattcaagg
 361 tccagggaat cctcagtgct tctctctccg aactgaacag gggacggtgc tcattggttc
 421 agaacatggc ctgatagaag tagaccctgt ctcaagagaa gtgaaaaatg aagtttcttt
 481 ggtggcagaa ggctttctcc cagaggatgg aagtggccgc attgttggtg ttcaggactt
 541 gctggatcag gagtctgtgt gtgtggccac agcctctgga gacgtcatac tctgcagtct
 601 cagcacacaa cagctggagt gtgttgggag tgtagccagt ggtatctctg ttatgagttg
 661 gagtcctgac caagagctgg tgcttcttgc cacaggtcaa cagaccctga ttatgatgac
 721 aaaagatttt gagccaatcc tggagcagca gatccatcag gatgattttg gtgaaagcaa
 781 gtttatcact gttggatggg gtaggaagga gacacagttc catggatcag aaggcagaca
 841 agcagctttt cagatgcaaa tgcatgagtc tgctttgccc tgggatgacc atagaccaca
 901 agttacctgg cggggggatg gacagttttt tgctgtgagt gttgtttgcc cagaaacagg
 961 ggctcggaag gtcagagtgt ggaaccgaga gtttgctttg cagtcaacca gtgagcctgt
1021 ggcaggactg ggaccagccc tggcttggaa accctcaggc agtttgattg catctacaca
1081 agataaaccc aaccagcagg atattgtgtt ttttgagaaa aatggactcc ttcatggaca
1141 ctttacactt cccttcctta aagatgaggt taaggtaaat gacttgctct ggaatgcaga
1201 ttcctctgtg cttgcagtct ggctggaaga ccttcagaga gaagaaagct ccattccgaa
1261 aacctgtgtt cagctctgga ctgttggaaa ctatcactgg tatctcaagc aaagtttatc
1321 cttcagcacc tgtgggaaga gcaagattgt gtctctgatg tgggaccctg tgaccccata
1381 ccggctgcat gttctctgtc agggctggca ttacctcgcc tatgattggc actggacgac
1441 tgaccggagc gtgggagata attcaagtga cttgtccaat gtggctgtca ttgatggaaa
1501 cagggtgttg gtgacagtct tccggcagac tgtggttccg cctcccatgt gcacctacca
1561 actgctgttc ccacaccctg tgaatcaagt cacattctta gcacaccctc aaaagagtaa
1621 tgaccttgct gttctagatg ccagtaacca gatttctgtt tataaatgtg gtgattgtcc
1681 aagtgctgac cctacagtga aactgggagc tgtgggtgga agtggattta aagtttgcct
1741 tagaactcct catttggaaa agagatacaa aatccagttt gagaataatg aagatcaaga
1801 tgtaaacccg ctgaaactag gccttctcac ttggattgaa gaagacgtct tcctggctgt
1861 aagccacagt gagttcagcc cccggtctgt cattcaccat ttgactgcag cttcttctga
1921 gatggatgaa gagcatggac agctcaatgt cagttcatct gcagcggtgg atggggtcat
1981 aatcagtcta tgttgcaatt ccaagaccaa gtcagtagta ttacagctgg ctgatggcca
2041 gatatttaag taccttttggg agtcaccttc tctggctatt aaaccatgga agaactctgg
2101 tggatttcct gttcggtttc cttatccatg cacccagacc gaattggcca tgattggaga
2161 agaggaatgt gtccttggtc tgactgacag gtgtcgcttt ttcatcaatg acattgaggt
2221 tgcgtcaaat atcacgtcat ttgcagtata tgatgagttt ttattgttga caacccattc
2281 ccatacctgc cagtgttttt gcctgaggga tgcttcattt aaaacattac aggccggcct
2341 gagcagcaat catgtgtccc atggggaagt tctgcggaaa gtggagaggg gttcacggat
2401 tgtcactgtt gtgccccagg acacaaagct tgtattacag atgccaaggg gaaacttaga
2461 agttgttcat catcgagccc tggttttagc tcagattcgg aagtggttgg acaaacttat
2521 gtttaaagag gcatttgaat gcatgagaaa gctgagaatc aatctcaatc tgatttatga
2581 tcataaccct aaggtgtttc ttggaaatgt ggaaaccttc attaaacaga tagattctgt
2641 gaatcatatt aacttgtttt ttacagaatt gaaagaagaa gatgtcacga agaccatgta
2701 ccctgcacca gttaccagca gtgtctacct gtccagggat cctgacggga ataaaataga
2761 ccttgtctgc gatgctatga gagcagtcat ggagagcata aatcctcata aatactgcct
2821 atccatactt acatctcatg taaagaagac aaccccagaa ctggaaattg tactgcaaaa
2881 agtacacgag cttcaaggaa atgctccctc tgatcctgat gctgtgagtg ctgaagaggc
2941 cttgaaatat ttgctgcatc tggtagatgt taatgaatta tatgatcatt ctcttggcac
3001 ctatgacttt gatttggtcc tcatggtagc tgagaagtca cagaaggatc ccaaagaata
3061 tcttccattt cttaatacac ttaagaaaat ggaaactaat tatcagcggt ttactataga
3121 caaatacttg aaacgatatg aaaaagccat tggccacctc agcaaatgtg gacctgagta
```

Figure 7

Continued

```
3181 cttcccagaa tgcttaaact tgataaaaga taaaaacttg tataacgaag ctctgaagtt
3241 atattcacca agctcacaac agtaccagga tatcagcatt gcttatgggg agcacctgat
3301 gcaggagcac atgtatgagc cagcggggct catgtttgcc cgttgcggtg cccacgagaa
3361 agctctctca gcctttctca catgtggcaa ctggaagcaa gccctctgtg tggcagccca
3421 gcttaacttt accaaagacc agctggtggg cctcggcaga actctggcag gaaagctggt
3481 tgagcagagg aagcacattg atgcggccat ggttttggaa gagagtgccc aggattatga
3541 agaagctgtg ctcttgctgt tagaaggagc tgcctgggaa gaagctttga ggctggtata
3601 caaatataac agactggata ttatagaaac caacgtaaag ccttccattt tagaagccca
3661 gaaaaattat atggcatttc tggactctca gacagccaca ttcagtcgcc acaagaaacg
3721 tttattggta gttcgagagc tcaaggagca agcccagcag gcaggtctgg atgatgaggt
3781 accccacggg caagagtcag acctctttct tgaaactagc agtgtcgtga gtggcagtga
3841 gatgagtggc aaatactccc atagtaactc caggatatca gcgagatcat ccaagaatcg
3901 ccgaaaagcg gagcggaaga agcacagcct caaagaaggc agtccgctgg aggacctggc
3961 cctcctggag gcactgagtg aagtggtgca gaacactgaa aacctgaaag atgaagtata
4021 ccatatttta aaggtactct ttctctttga gtttgatgaa caaggaaggg aattacagaa
4081 ggccttgaa gatacgctgc agttgatgga aaggtcactt ccagaaattt ggactcttac
41 ttaccagcag aattcagcta ccccggttct aggtcccaat tctactgcaa atagtatcat
ggcatcttat cagcaacaga agacttcggt tcctgttctt gatgctgagc tttttatacc
accaaagatc aacagaagaa cccagtggaa gctgagcctg ctagactgag tgactgcagt
taggagggat ccgacagaga agaccatttc cactcattcc tgttgtccta ccaccccttg
ctctttgagg gctggctatt gagaactgga aagagtaaaa tgataactta ccttagcatt
4441 gccaagaact tcagcagaca acaagcaatt ctatttattt tatgttgtgt atacatcttg
4501 atcattagca agacattaag ctttaaccat tatggcacca ttttgtgaga atgattgttc
4561 tttcacttgg gctgtttgag agcataatta tggtaatcat gagattaatg tttcatgatt
4621 tctacctcca aagtgtgaag acaagtaaaa caatgtttct aaattgtctt attttgttgg
4681 cggagaagat tacaatggct attagtgcta catttggtca aatgtaatca cttaaatagc
4741 ttcttgtcac cttaaactaa agcagaataa aaagtatcct ttgaaattat aagccctcct
4801 ttgctgacag ctattatttt gtaacatctt accaggtcat gtgctttcag ttataactgg
4861 gctgagcctc ctataattac aatgtctata gggactgttt tactgcctgt gtatttttctg
4921 ctagagagtt agcaatgtta gagctagaac agattagaat ttctaaacag tatcatgcac
4981 agttggtgtg agtgatcagt gtgcattgta tggcatgcat ggttgtgaat tattctctgt
5041 tctccaaata ctgtttcttt aactcagata tttttgttag tgtctaggcc acttcattta
5101 tttttcgtca tggtacttta ctgacttctc tttattcaat tctccacgcc ctcaccaaaa
5161 aaaactgtct caaaatgaga atatttttat tcttcatggt ggtgctagaa aacgccccac
5221 ttcattctga ttaaaaaatt cttccatgtt tttaaatatc agaaccagac ctttcttact
5281 gtgtatctta gcccatttgt gtctctataa caacaaccag ctttcaaagg aactaataga
5341 gtgaaaactc actcattacc acgaggatgg cacaagcgat tcacgtagga tctgcccctg
5401 tgaccaaaac acctcccatt gggccacact tccaacactg gtgatcacat ttcaacatga
5461 ggtttaggga aacaaatgcc taaactacag cactgtacat aaactaacag gaaatgctgc
5521 ttttgatcct caaagaagtg atatagccaa aattgtaatt taagaagcct ttgtcagtat
5581 agcaagatgt taactataga atcaatctag gagtattcac tgtaaaattc aacttttctg
5641 tatgtttgaa cattttcaca atctcatagg agtttttaaa aagaagagaa agaagatata
5701 ctttgctttg gagaaatcta ctttttgact tacatgggtt tgctgtaatt aagtgcccaa
5761 tattgaaagg ctgcaagtac tttgtaatca ctctttggca tgggtaaata agcatggtaa
5821 cttatattga aatatagtgc tcttgctttg gataactgta aagggaccca tgctgataga
5881 ctggaaatag aagtaaatgt gtttattgaa aaaaaaaaaa aaaa
```

FIGURE 8

```
   1 mrnlklfrtl efrdiqgpgn pqcfslrteq gtvligsehg lievdpvsre vknevslvae
  61 gflpedgsgr ivgvqdlldq esvcvatasg dvilcslstq qlecvgsvas gisvmswspd
 121 qelvllatgq qtlimmtkdf epileqqihq ddfgeskfit vgwgrketqf hgsegrqaaf
 181 qmqmhesalp wddhrpqvtw rgdgqffavs vvcpetgark vrvwnrefal qstsepvagl
 241 gpalawkpsg sliastqdkp nqqdivffek ngllhghftl pflkdevkvn dllwnadssv
 301 lavrledlqr ekssipktcv qlwtvgnyhw ylkqslsfst cgkskivslm wdpvtpyrlh
 361 vlcqgwhyla ydwhwttdrs vgdnssdlsn vavidgnrvl vtvfrqtvvp ppmctyqllf
 421 phpvnqvtfl ahpqksndla vldasnqisv ykcgdcpsad ptvklgavgg sgfkvclrtp
 481 hlekrykiqf ennedqdvnp lklglltwie edvflavshs efsprsvihh ltaassemde
 541 ehgqlnvsss aavdgviisl ccnsktksvv lqladgqifk ylwespslai kpwknsggfp
 601 vrfpypctqt elamigeeec vlgltdrcrf findievasn itsfavydef llltthshtc
 661 qcfclrdasf ktlqaglssn hvshgevlrk vergsrivtv vpqdtklvlq mprgnlevvh
 721 hralvlaqir kwldklmfke afecmrklri nlnpiydhnp kvflgnvetf ikqidsvnhi
 781 nlfftelkee dvtktmypap vtssvylsrd pdgnkidlvc damravmesi nphkyclsil
 841 tshvkkttpe leivlqkvhe lqgnapsdpd avsaeealky llhlvdvnel ydhslgtydf
 901 dlvlmvaeks qkdpkeylpf lntlkkmetn yqrftidkyl kryekaighl skcgpeyfpe
 961 clnlikdknl ynealklysp ssqqyqdisi aygehlmqeh myepaglmfa rcgahekals
1021 afltcgnwkq alcvaaqlnf tkdqlvglgr tlagklveqr khidaamvle esaqdyeeav
1081 llllegaawe ealrlvykyn rldiietnvk psileaqkny mafldsqtat fsrhkkrllv
1141 vrelkeqaqq aglddevphg qesdlfsets svvsgsemsg kyshsnsris arsskrrrka
1201 erkkhslkeg spledlalle alsevvqnte nlkdevyhil kvlflfefde qgrelqkafe
1261 dtlqlmersl peiwtltyqq nsatpvlgpn stansimasy qqqktsvpvl daelfippki
1321 nrrtqwklsl ld
```

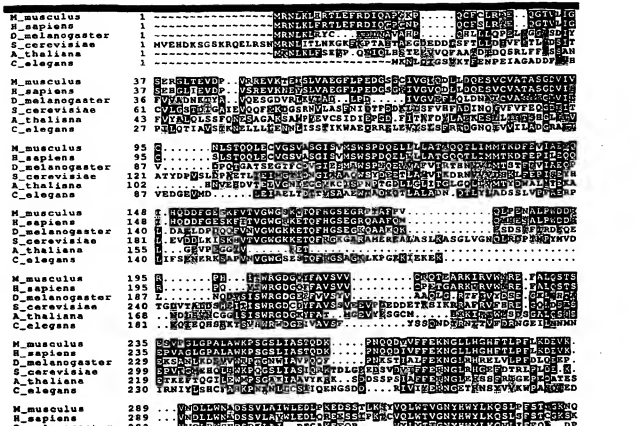


FIG. 1. Comparison of the amino acid sequence of Ikap across several species. Alignment of the amino acid sequence of Ikap (M_musculus) with that of *Homo sapiens* (H_sapiens), *Drosophila melanogaster* (D_melanogaster), *Saccharomyces cerevisiae* (S_cerevisiae), *Arabidopsis thaliana* (A_thaliana), and *Caenorhabditis elegans* (C_elegans). Black boxes indicate identical AA, while conserved AA residues are shown in gray. Asterisk (*) at AA position 606 for mouse and human proteins indicates the location of the heterozygous R606P mutation found in only 4 FD patients. Sequence alignments were made using Pileup and Boxshade commands from GCG Wisconsin Package V.9.0 (Madison, WI).

Figure 9

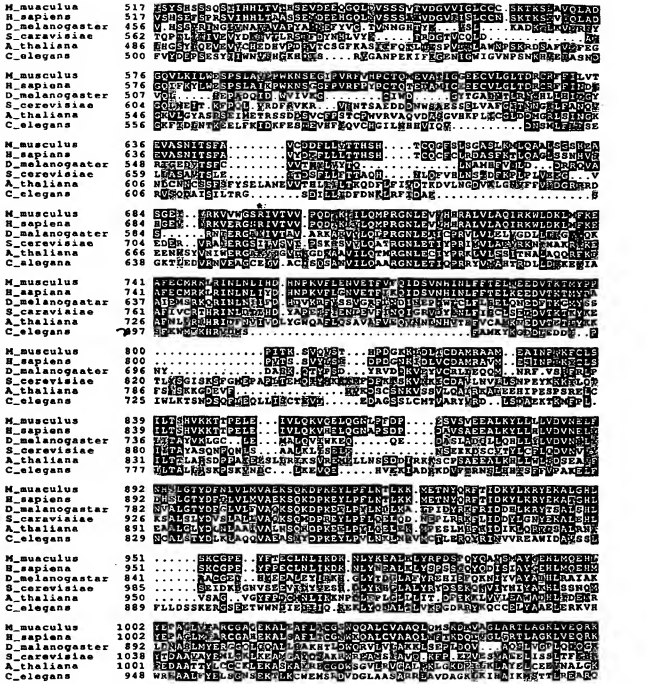


Figure 9
Continued

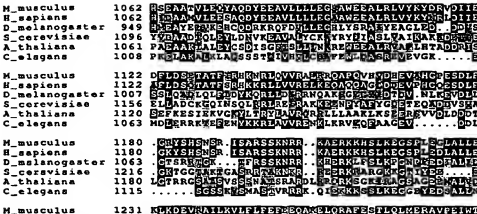


Figure 9
Continued

TABLE 2. COMPARISON OF THE NOVEL MOUSE *Ikbkap* GENE WITH MULTIPLE SPECIES HOMOLOGS

Species	*Gene name*	*No. of amino acids*	*Molecular weight (kDa)*	*% aa identity with M.m.*	*GenBank Accession No.*
Mus musculus (M.m.)	*Ikbkap*	1332	149.11	—	AF367244
Homo sapiens	*IKBKAP*	1332	149.11	80	AF153419
Drosophila melanogaster	*CG10535*	1213	138.21	32	AAF54670
Saccharomyces cerevisiae	*Elp1/Iki3p*	1349	152.99	29	AAB67278
Arabidopsis thaliana	Unknown	1308	146.63	27	BAB08695
Caenorhabditis elegans	Unknown	1177	134.80	24	AAF60430

Figure 10

TABLE 1. MOUSE *Ikbkap* EXON AND INTRON BOUNDARIES

Exon	*Acceptor site*	*Donor site*	*Size (bp)*	*cDNA position*
1		AGgtgagcatlcgcccg	129	1..129[a]
2	ttttttttccctcagAA	AAgtaggtcactgatgc	163	130..292[b]
3	tatgctttgtgaaagGT	AGgtaggtgtaaggcct	153	293..445
4	tttctctgatgcagCT	AGgtaagctttgcactg*	82	446..527
5	acatgaactcctaagCT	AGgtaagcgttttttgg	81	528..608
6	ctlgaaaaactgtagGC	TGgtaaggcgggatgat	86	609..694
7	ggtgtctctcttcagCC	TGgtgtctctcttcagc*	97	695..791
8	ctacctccttgcagAG	AAgtgagtgagcataaa*	91	792..882
9	aggttctgctttcagAC	AGgtagggggtcagagtt	124	883..1006
10	ttttgtcccttaccagGT	TGgtatgacagcttgtg	94	1007..1100
11	tccctccacacacagTC	AAgtaagttgcctgcgaa	231	1101..1331
12	ctttttcattgtgtagAC	TGgtaagtggaagcagg	165	1332..1496
13	ttttgtgtttctagGT	TCgtaagttcctaaata	100	1497..1596
14	ctaatatttgaacagGA	AGgtatcatggttcatc	189	1597..1785
15	ttttttttgctttagTT	GGgtggagatcagagtt	107	1786..1892
16	ttaatcttacaacagAG	AGgtgaatagacacggc	104	1893..1996
17	ttcatttctttgcagGA	AGgtatgtaggcttggt	54	1997..2050
18	tctlgcctgttgcagGT	AAgtaagctctcctata	106	2051..2156
19	cactggtatttttagTG	AGgtaagctgactcttc*	116	2157..2272
20	gggttttatttgagAT	AAgtaagtatttattct*	74	2273..2346
21	ttcctgtcctcacagAC	AGgtcacattgcgtct	79	2347..2425
22	tacttttcttgatagGT	AGgtaagtatttgata*	80	2426..2505
23	tactgtggttctagGG	AAgtgggtgctgtgtgt	138	2506..2643
24	cactactacctcagGT	AGgtagagacctgcgcg*	86	2644..2729
25	cttaaactccaacagGA	AGgtatgtggagtgag*	149	2730..2878
26	aacttttttcctaggGA	TGgtaagggttttttt	124	2879..3002
27	ttttttttttttcagGA	AGgtatgtggtgggtta*	98	3003..3100
28	cgtctcttgtcacagGC	AGgtaagcagggccatt	202	3101..3302
29	ttgctgtctttcagGA	AGgtgagctcctcccgg	62	3303..3364
30	ctcttccctgtcagGA	TGgtaaggaagctctga	63	3365..3427
31	tttctccctcctagGT	AGgtgaggattacattt*	61	3428..3488
32	attatgcatcctcagCC	GGgtgagtgcctccaaa*	114	3489..3602
33	gttcatcttctctagAT	GCgtacgtacgagacct*	112	3603..3714
34	tgtaatttctgacagGA	AGgtatggcttcagtgc	128	3715..3842
35	ccatttctttctctagAT	CGgtaagcttcctcaga	155	3843..3997
36	ctgtttctgcttagGT	CGgtgtactgctcgttc	76	3998..4073
37	cattcttgcttccagAT		709	4074..4799[c]

Figure 11

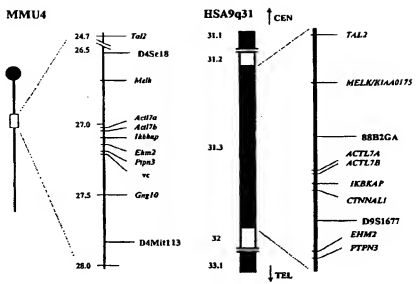


Figure 12